

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:27:13 ; Search time 14 Seconds

(Without alignments)
1653.128 Million cell updates/sec

Title: US-10-037-311a-1

Predicted score: 3004

Sequence: 1 MDQNSYRRSSPIRTTGS.....GTLVPHVRCEDISWGLKLV 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3004	100.0	558	1	095wh5 arabidopsis
2	1776	59.1	565	1	FUT1_ARATH
3	1706.5	56.8	526	1	FUT1_PEA
4	1679	55.9	539	1	FUT2_ARATH
5	1613.5	53.7	500	1	FUT8_ARATH
6	1583	52.7	537	1	FUT6_ARATH
7	1528	50.9	503	1	FUT4_ARATH
8	1492	49.7	533	1	FUT5_ARATH
9	1411.5	47.0	440	1	FUT10_ARATH
10	1302	43.3	435	1	FUT9_ARATH
11	1113.5	37.1	493	1	FUT3_ARATH
12	110.5	3.7	322	1	NOD2_ARHNS
13	97	3.2	594	1	UVRG_HELPJ
14	97	3.2	1224	1	MDRB_HELPJ
15	96	3.2	594	1	UVRG_HELPJ
16	94	3.1	1333	1	CC25_CANAL
17	94	3.1	1376	1	VG12_CVMA5
18	93.5	3.1	807	1	INFE_ECOLI
19	93	3.1	455	1	PR17_YEAST
20	93	3.1	673	1	COAT_PAVBO
21	92.5	3.1	993	1	NISB_LACIA
22	92	3.1	395	1	NEUA_STRAG
23	92	3.1	545	1	REC8_RICCN
24	92	3.1	1376	1	VG12_CVMA
25	92	3.1	1581	1	ANOL_PNECA
26	91	3.0	536	1	CATA_DEIRA
27	90.5	3.0	415	1	SACB_PSEEG
28	90.5	3.0	492	1	CAT3_ARATH
29	90.5	3.0	847	1	DNL1_CAEEL
30	90.5	3.0	848	1	AHR_HUMAN
31	90.5	3.0	1612	1	RRPO_ORSVS
32	90	3.0	561	1	V423_MYCPN
33	90	3.0	1235	1	VG12_CVMA5

34	90	3.0	3080	1	POLG_ZMYWC	P18479 z genome po
35	90	3.0	3746	1	ACV7_PENCH	P19787 penicillium
36	90	3.0	3791	1	PATC_PENCH	P26046 penicillium
37	89.5	3.0	1286	1	NI59_YEAST	P18502 drosophila
38	89.5	3.0	1460	1	GAL7_SCHPO	P40477 saccharomyc
39	89	3.0	369	1	POLG_ZYMS	O36979 z genome po
40	89	3.0	3083	1	RIP2_MOUSE	P97433 mus musculu
41	88.5	2.9	1693	1	RRPL_SYNY	P31332 sonchus yel
42	88.5	2.9	2116	1	RRPL_SYNY	P22372 claviclops p
43	88	2.9	936	1	RRPO_HUMAN	O60282 homo sapien
44	88	2.9	957	1	RRPO_HUMAN	P11224 murine coro
45	88	2.9	1324	1	VG12_CVMA5	

ALIGNMENTS

RESULT 1
FUT1_ARATH STANDARD: PRT: 558 AA.
AC 095WH5: 081052;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactoside 2-L-fucosyltransferase (EC 2.4.1.69) (Xyloglucan alpha-(1,2)-fucosyltransferase) (Atrfuf1).
GN FUT1 OR FRT1 OR AT2G03220 OR T18E12.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=cv. Columbia.
RC MEDLINE=99301928; PubMed=10373113;
RA Perrin R.M., Derocher A.E., Bar-Peled M., Zeng W., Norambuena L.,
RA Orellana A., Ralshel N.V., Keegstra K.;
RT "Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall
RT biosynthesis.";
RT Science 284:1976-1979(1999).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia.
RC MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umeyam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Niceman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
[3]
RP FUNCTION, AND TISSUE SPECIFICITY.
RC MEDLINE=21608393; PubMed=11743104;
RA Sarria R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,
RA Keegstra K., Ralshel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
RT xyloglucan fucosyltransferase.";
RL Plant Physiol. 127:1595-1606(2001).
CC -i- FUNCTION: Involved in cell wall biosynthesis. Adds the terminal
CC fucosyl residue on xyloglucan side chains.
CC -i- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +
CC alpha-L-fucosyl-1,2-beta-D-galactosyl-R.
CC -i- PATHWAY: Glycosylation.
CC -i- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC in trans cisternae of Golgi (By similarity).
CC -i- TISSUE SPECIFICITY: Expressed in roots, stems, leaves, flowers,
CC siliques and seedlings.

```
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF154111: AAD41092.1: -
CC EMBL: AC005313: AAC34480.1: -
CC InterPro: IPR004938: XG_Ftase.
CC Pfam: PF03254: XG_Ftase; 1.
CC Transferase: Glycosyltransferase; Transmembrane; Glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC DOMAIN 1 43 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 44 64 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC DOMAIN 65 558 LUMENAL, CATALYTIC (POTENTIAL).
CC CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 270 270 V -> F (IN REF. 1).
CC SEQUENCE 556 AA: 63452 MW: 1BF0E372231A8ADF CRC64:
Query Match 100.0%; Score 3004; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 3.5e-234;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOONSTRARRRRSIRRTTGGSKSVNSSELLQMKYLSGTMKLTTRTTTCLIVSVLAVSM 60
DB 1 MOONSTRARRRRSIRRTTGGSKSVNSSELLQMKYLSGTMKLTTRTTTCLIVSVLAVSM 60
QY 61 IFHQHPSDSNRIMGFAEARVLDAGVFPNTNNSDKLGGILASGFDESDCSRQSVHY 120
DB 61 IFHQHPSDSNRIMGFAEARVLDAGVFPNTNNSDKLGGILASGFDESDCSRQSVHY 120
QY 121 RKPSPYKSSSYISKLRYNEKILHKRCGPCTESYKALQOLDQEHIDGECRYVWISFS 180
DB 121 RKPSPYKSSSYISKLRYNEKILHKRCGPCTESYKALQOLDQEHIDGECRYVWISFS 180
QY 121 RKPSPYKSSSYISKLRYNEKILHKRCGPCTESYKALQOLDQEHIDGECRYVWISFS 180
DB 121 RKPSPYKSSSYISKLRYNEKILHKRCGPCTESYKALQOLDQEHIDGECRYVWISFS 180
QY 181 GIGNNLSIASVFLALTDRLVLDVRGKMDLCEPFLGSMILPLDPMPTDOFGLN 240
DB 181 GIGNNLSIASVFLALTDRLVLDVRGKMDLCEPFLGSMILPLDPMPTDOFGLN 240
QY 241 QESSRCYGYWAKNOYIDTEGTLSHLYLHVDHYGDHDMFCEGQPTLGKVPMLIVTD 300
DB 241 QESSRCYGYWAKNOYIDTEGTLSHLYLHVDHYGDHDMFCEGQPTLGKVPMLIVTD 300
QY 301 NYFVPSMLILGFDDELAKLFPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
DB 301 NYFVPSMLILGFDDELAKLFPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
QY 301 NYFVPSMLILGFDDELAKLFPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
DB 301 NYFVPSMLILGFDDELAKLFPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
QY 361 IGIOVRVDEDEPGPOHMDIOISCTQEKILPEYDVLVESSRHVNTPKHAAVLVTSINA 420
DB 361 IGIOVRVDEDEPGPOHMDIOISCTQEKILPEYDVLVESSRHVNTPKHAAVLVTSINA 420
QY 421 GYAENLKSMEWEPYPTSTGEIIGVHOPSOEGYQOTEKMHNGKALAEWLLSLTDNLVISA 480
DB 421 GYAENLKSMEWEPYPTSTGEIIGVHOPSOEGYQOTEKMHNGKALAEWLLSLTDNLVISA 480
QY 481 WSTGTYVAQGIAGGLKFWLLYPENKTTTPDPSCGRAMSKPECFHSPPYDCAKTGIDTGT 540
DB 481 WSTGTYVAQGIAGGLKFWLLYPENKTTTPDPSCGRAMSKPECFHSPPYDCAKTGIDTGT 540
QY 541 LVPVHRCEDISWGLKTV 558
DB 541 LVPVHRCEDISWGLKTV 558
RESULT 2
FUT1_PEA STANDARD; PRT; 565 AA.
ID FUT1_PEA
```

```
AC Q9W501:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactoside 2-L-fucosyltransferase (EC 2.4.1.69) (Xyloglucan alpha-
DE (1,2)-fucosyltransferase) (PsfT1).
GN PsfT1.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=cy Alaska;
RX MEDLINE=20270214; PubMed=10747946;
RA Falk A., Bar-Peled M., Derocher A.E., Zeng W., Perrin R.M.,
RA Wilkerson C., Raikhel N.V., Keegstra K.;
RT "Biochemical characterization and molecular cloning of an
RT alpha-1,2-fucosyltransferase that catalyzes the last step of cell wall
RT xyloglucan biosynthesis in pea."
RT J. Biol. Chem. 275:15082-15089(2000).
CC -!- FUNCTION: Involved in cell wall biosynthesis. Adds the terminal
CC fucosyl residue on xyloglucan side chains.
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +
CC alpha-L-fucosyl-1,2-beta-D-galactosyl-R.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF223643: AAF62886.1: -
CC InterPro: IPR004938: XG_Ftase.
CC Pfam: PF03254: XG_Ftase; 1.
CC Transferase: Glycosyltransferase; Transmembrane; Glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC DOMAIN 1 43 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 44 64 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC DOMAIN 65 565 LUMENAL, CATALYTIC (POTENTIAL).
CC CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 565 AA: 64029 MW: 5E6AAABED20594B CRC64:
Query Match 59.1%; Score 1776; DB 1; Length 565;
Best Local Similarity 62.6%; Pred. No. 2.8e-135;
Matches 330; Conservative 69; Mismatches 116; Indels 12; Gaps 4;
QY 39 MKLRTTTCCLIVSVLVASMTFHHQSPD---SNRMGFAEARVLDAGVFPNTNNSD 95
DB 42 MRVAFVVSFEMFLSVLFSVLRDPPSDAIAISSTTLTFLQNLGSLSDSDFSVELLN 100
QY 96 KILGGLASGFDESDCSRQSVHYRKPSPYKSSSYISKLRYNEKILHKRCGPCTESYK 155
DB 101 KILGGLADGFDEKSCSRQSAIFGKLSGKSSSYISKLRYNEKILHKRCGPCTESYK 160
QY 156 ALKOLDQEHIDGECRYVWISFSFGNRIILASVFLALTDRLVLDVRGKMDLPLF 215
DB 161 TYKELSGQGSSEYDCKYVWISFSGLGNRIILVLSFLALTDRLVLDVRGKMDLPLF 220
QY 216 CEPFLGSMILPLDPMPTDOFGLNDESSRCYGYWAKNOYIDTEGTLSHLYLHVDYGD 275
DB 221 CEPFPDASWFWPPDFPLNSHLNNEGNSNCHGKIILKTSITNSYPSFVYLLHLDYDD 280
```


QY	276	HDKHEFCEGDOTFLGKXYPMLIVKTDNYFVPSLMLIDFEDDELKLPPOKQATVFNHILGRYL	335
Db	281	HDKLEFDEDEQLQONVPLIMKMDNFISLTSLSFDEOLDPEKKKEKVFHFGRLT	340
QY	336	FHPNQVWGILTRYEAYVLSHADKIGIQVRVDEDEPGRPOHMOOSSCQTEKLLPEY	395
Db	341	LHPNNWVGGLVRYRYDAYLAKVBERIGIQIRVFDTPDGPQOHNLDQVLACTLKESILPEY	400
QY	396	DLTVERSHNV----TPKHXAVLTVSLNAGYAEHLKSMTHREYPTSTGEILVGHQPSDEGT	451
Db	401	N-----RCQNLINSSSGTGRSKAVLITSLSSGYEKKVRMWEFPLETGELVAGVGYOPSHCY	456
QY	452	QOOTEKKMHNGKALAEWYLLSLTDNLVYLSASGTGYAAGIGLGLKPMILVPERNRTPDSS	511
Db	457	QOQOKOFHNGKAMAEWYLLSLTDVLVYLSMSSTGYAAGIGLGLKPMILVPERNRTPNRP	516
QY	512	CGRAMSMERCHSPFPFDCAKAGTIDTGILVPHVRHCEDISGKLKLV	558
Db	517	CGRAMSMERCHAPFPYDCAKAKCTDGVALLVPHVRHCEDISGWLKLV	563

RESULT 3				
FUT7_ARATH	STANDARD;	PTT;	526	AA.
ID	FUT7_ARATH			
AC	09X181; O9LMF2;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Probable fucosyltransferase 7 (EC 2.4.1.-) (AtFUT7).			
GN	FUT7 OR AT1G14070 OR F7A19.15 OR F16A14.19 OR F16A14.28.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; streptophyta; Embryophyta; Tracheophyta;			
OC	Eumetazoa; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	spermatids II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RC	MEDLINE=21016719; PubMed=11130712;			
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,			
RA	White O., Alonso J., Altif L., Ararajo R., Bowman C.L., Brooks S.Y.,			
RA	Buehler E., Chao Q., Chen H., Chau K.F., Chin C.W.,			
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,			
RA	Dunn P., Etlip P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,			
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,			
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,			
RA	Kim C.J., Koo H.L., Kremetskaia I., Kutz D.B., Kwan A., Lam B.,			
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,			
RA	Lih X., Liu S.X., Liu Z.A., Lueros J.S., Malt R., Marziani A.,			
RA	Miltschert J., Miranda M., Nguyen M., Niernann W.C., Osborne B.I.,			
RA	Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,			
RA	Saiano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,			
RA	Sun H., Tallon L.J., Tamunga G., Tortum M.J., Town C.D.,			
RA	Utecherak T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,			
RA	Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;			
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis			
RL	thaliana.";			
RL	Nature 408:816-820(2000).			
RN	[2]			
RP	IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY			
RP	MEDLINE=21608399; PubMed=11743104;			
RX	Sarris R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,			
RA	Keegstra K., Rakheh N.V.;			
RT	"Characterization of a family of Arabidopsis genes related to			
RT	xyloglucan fucosyltransferase1.";			
RL	Plant Physiol. 127:1595-1606(2001).			
CC	-1- FUNCTION: May be involved in cell wall biosynthesis. May			
CC	act as a fucosyltransferase.			
CC	-1- PATHWAY: Glycosylation.			
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound			
CC	form in trans cisternae of Golgi (By similarity).			
CC	-1- TISSUE SPECIFICITY: Expressed in roots, leaves, stems and			

CC	seedlings.	
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.	
CC	-I- CAUTION: Ref:1 (AAE79408) sequence differs from that shown due to	
CC	exonous gene model prediction.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb.ch/announce/	
CC	or send an email to license@isb.slb.ch).	
CC	-----	
DR	EMBL: AC007576; AAD39292.1; -	
DR	EMBL: AC068197; AAE79408.1; ALT_SEQ.	
DR	InterPro: IPR004938; XG_F1ase.	
DR	Pfam: PF03254; XG_F1ase: 1.	
KW	Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;	
KW	Signal-anchor; Golgi stack; Cell wall.	
FT	DOMAIN	1 4
FT	TRANSMEM	5 25
FT	(POTENTIAL).	
FT	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)	
FT	(POTENTIAL).	
FT	LUMINAL, CATALYTIC (POTENTIAL).	
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	211 211
FT	CARBOHYD	215 215
FT	CARBOHYD	363 363
FT	SEQUENCE	526 AA; 60537 MW; 863B447551025FCD CAC64;
SO		

Query Match	56.8%	Score 1706.5;	DB 1;	Length 526;
Best Local Similarly	62.2%	Pred. No. 1e-129;		
Matches 332; Conservative	72;	Mismatches 97;	Indels 33;	Gaps 9;

[illegible]


```

AC 095J94: 15-JUN-2002 (Rel. 41, Created)
AD 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable fucosyltransferase 5 (EC 2.4.1.-) (AtFUT5).
GN FUM5 OR ATG215370 OR F26H6.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A..
RX Medline=21608393; Pubmed=11743104;
RA Sarría R., Wagner T.A., O'Neill M.A., Paik A., Wilkerson C.G.,
RA Keegstra K., Raikhel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
RL xyloglucan fucosyltransferase.";
RL Plant Physiol. 127:1595-1606(2001).
RN [2]
RP SEQUENCE FROM N.A..
RC STRAIN=cv. Columbia;
RX Medline=20083487; Pubmed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H. L.,
RA Motif K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Niernm W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana.";
RL Nature 402:761-768(1999).
CC -I- FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC -I- PATHWAY: Glycosylation.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -I- TISSUE SPECIFICITY: Expressed in roots, leaves, flowers and
CC siliques.
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF417475; AAL50624.1; ALT_INIT.
DR EMBL: AC006920; AAD22287.1; -.
DR InterPro: IPR004938; XC_FTase.
DR Pfam: PF03254; XG_FTase; 1.
KW Transferase; Glycosyltransferase; Transmembrane; glycoprotein;
KW Signal-anchor; Golgi stack; Cell wall.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 35 533 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 533 AA; 61173 MW; 63CE63AFAD56E43 CRC64;
Query Match 49.7%; Score 1492; DB 1; Length 533;
Best Local Similarity 54.6%; Pred. No. 1,9e-112;
Matches 295; Conservative 90; Mismatches 133; Indels 22; Gaps 9;

```

Db	1	MYKFOISG - KIYKLTGLKKAKVLIASFSSLLF-----ILSTSNPNKKLIDATT	49
Qy	86	FPNV - ININSDKLTGLGSLDEDESCLEXYOVH - YRKSPYKFSYSLISKLRNVEKL	142
Db	50	KVDIKETEKRPDKLIGGLTLADPDESGCLSRHKHYFLYRKPSRYKSEYLSKLSREYML	109
Qy	143	HKRCGPGETESYKAKKOLDOEHT - DGDGECKYVWVISFSGNRIILSLASVFLYALLTR	201
Db	110	HKRCGPDEYKKEAIKELSLRDSASESNGECRYIWMVAGYGLGNRLTLTASVFLYALLTER	169
Qy	202	VLLVDRGCDMDLCEPFFLGSLWLLPLDFPMTQFD--GINOSSKCYGMKNQYIDE	259
Db	170	IILVDNRKVDYDLCEPFPGLSWLLPLDFPMLVITYAMGYNKPEPCYGMSEKHSINST	229
Qy	260	GLTSLHLYLHVODYGDHDKMFCFGDQDTFGKAPWMLVTKDNEFVPSLMLPCFDDELNK	319
Db	230	SIRPHLYVHNLHSDRSDSKLFCYCKDQSLDKPMLVQANVYFVPSLWPNPFFQELVK	289
Qy	320	LEPKATVFNHLLGKRYLFHPNTQWGLVTRYEAVLSHADEKIGIYRVDEDEDPQHW	379
Db	290	LEPKETEFVHLLARYLEHPNNEVMDVTDYVHMLSKADERLGIQLRVGKPDGRKHVI	349
Qy	380	DOISSCTQKEKLEPVDTLYERSHNV - TRHKAVLYTSLNAGYAENLSMVEYPTSTG	438
Db	350	DOVISCQREKLELPEFATPEESKKNISKTRKLSVLAASLPEFSGNLTMFSKRPSSTG	409
Qy	439	EIIIGHOPSGOGYQOTKKHNGKALAEMLLSTDLVLSAMSTGAYVAGQGLKPMI	498
Db	410	EIVEYQDPSGRVQOTDKSHIDKALAEMLLSTDLVITSASTGCVYSISLGGKPMI	469
Qy	499	LYRENTTTPDPSGGRAMSEPCFSPFYDCAKAKTIDGTLVPHVHCEDISW - GLKL	557
Db	470	LYQPTNFTTPMPPCVRKSMPCYLTPPSSHGCEADMGTSKILPVRHCEDLIYGLKL	529
RESULT 9			
FU01_	ARATH	STANDARD;	PRT; 440 AA.
AC	Q95JP6;		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Putative fucosyltransferase 10 (EC 2.4.1.-) (AtFU10) (Fragment).		
GN	FU10 OR AT2615350 OR F26H6.13.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	Eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia;		
RC	MEDLINE=20083487; PubMed=10617197;		
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,		
RA	Fuji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,		
RA	Buell C.R., Keichum K.A., Lee J.J., Ronning C.M., Koo H.L.,		
RA	Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,		
RA	Tallon L.J., Gill J.E., Adams M.D., Garreta A.J., Cressy T.H.,		
RA	Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,		
RA	Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,		
RA	Venter J.C.;		
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis		
RT	thaliana."		
RL	Nature 402:761-768(1999).		
RN	[2]		
RP	IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY		
RP	MEDLINE=21608393; PubMed=11743104;		
RA	Sarría R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,		
RA	Keegstra K., Rakhel N.V.,		
RT	"Characterization of a family of Arabidopsis genes related to		
RT	xyloglucan fucosyltransferase1."		
RL	Plant Physiol. 127:1595-1606(2001).		

```

CC -!- FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC
CC -!- PATHWAY: Glycosylation.
CC
CC -!- SUBCELLULAR LOCATION: Golgi (potential).
CC
CC -!- TISSUE SPECIFICITY: Expressed in root, leaves, stems and
CC seedlings.
CC
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -!- CAUTION: AtFUT10 (as currently annotated) is lacking the first
CC exon and intron. It is therefore incomplete and we awaits cDNA
CC confirmation of the N-terminal extension.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC006920; AAD22289.1; -.
CC InterPro: IPR004938; XG_FTase.
CC Pfam: PF03254; XG_FTase; 1.
CC
CC KW Hypothetical protein; Transferase; Glycosyltransferase; Glycoprotein;
CC Golgi stack; Cell wall.
CC
CC FT NON_TER 1 1
CC FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SEQUENCE 440 AA; 50837 MW; 2469BD38934080 CRC64.
CC
CC Query Match 47.0%; Score 1411.5; DB 1; Length 440;
CC Best Local Similarity 59.5%; Pred. No. 4.4e-106;
CC Matches 261; Conservative 75; Mismatches 90; Indels 13; Gaps 5;
CC
CC QY 128 PSSYLSKRNANKKRGCPGTEYSKAKLKDDEH--DGDGECKYVWISFGGLGN 184
CC ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 2 PSEYLVSELRSEEMHKRGCPRTKATKTEKLSREYVASESNGECRYIWLARDGLGN 61
CC
CC QY 185 RLTLASVFLVALLDRLVLDVRGKMDLFCPEFLGMSWLPDPEPMD--QDFGLDNE 242
CC ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 62 RLTLASVFLVALLDRLVLDVRGKMDLFCPEFLGMSWLPDPEPMD--QDFGLDNE 121
CC
CC QY 243 SSRCTGVYMKNOVIDEGLSHLYLHLVHVDHDKMFECEGDTFTGKRWLYKTNDY 302
CC |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 122 YRCYGTMLNENIAINSTSIPRLYLHNHDSRDSKLFECQKDSFIDKVPWLIQTNAV 181
CC
CC QY 303 FVPSLWLVGDFDELNKLPKOKATVFNHILGRLEFHTPTQVWGLVTRYRYEATYSHADEKIG 362
CC |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 182 FVPSLWLVGDFDELNKLPKOKATVFNHILGRLEFHTPTQVWGLVTRYRYEATYSHADEKIG 241
CC
CC QY 363 IOVRVDEDDPGFPOHVMDOISSCTQKELLPVDPDLVRSRHVN--TPKIKAVLVTSIN 419
CC ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 242 IOVRVDEDDPGFPOHVMDOISSCTQKELLPVDPDLVRSRHVN--TPKIKAVLVTSIN 297
CC
CC QY 420 AGYAEKLSMTVEYPTSTGEITIGVHOPSQOEGYQOTEEKMHNKALAEWYLLSTDNLVTS 479
CC ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 298 PEYSVNLTMFLARSSSTGEILEVYQPSAEYQOQDKKSHQKALAEWYLLSTDNLVTS 357
CC
CC QY 480 AMSTGCVYVAGGGLKPMVLYRPPKRTTDPSCGRAMMECPHSPPYDDCAKATGIDTG 539
CC |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC 358 GMSITGCVYVAGGGLKPMVLYRPPKRTTDPSCGRAMMECPHSPPYDDCAKATGIDTG 357
CC
CC DB 540 TLVPHVRHCEIDSM--GLKL 557
CC ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
CC DB 418 KILPVRHCEIDSMYGLKL 436
CC
CC RESULT 10
CC FUT9_ARATH STANDARD; PRT; 435 AA.
CC AC 09X177;

```

```

DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable fucosyltransferase 9 (EC 2.4.1.-) (AtFUT9).
GN FUT9 OR AT1G14110 OR F7A19.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustrodi II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia.
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Paim C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altali H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen A., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huzar L.,
RA Hunter J.L., Jenkins J., Johnson Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.L.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Taitou L.J., Tumbunga G., Turlum M.V., Town C.D.,
RA Uteckack T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [2]
RP IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE AND TISSUE SPECIFICITY.
RX MEDLINE=21608393; PubMed=11743104;
RA Sarrila R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,
RA Keegstra K., Raikhel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
RT xyloglucan fucosyltransferase1."
RL Plant Physiol. 127:1595-1606(2001).
CC -!- FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC
CC -!- PATHWAY: Glycosylation.
CC
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC
CC -!- TISSUE SPECIFICITY: Expressed in leaves and stems.
CC
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC007576; AAD39295.1; -.
CC InterPro: IPR004938; XG_FTase.
CC Pfam: PF03254; XG_FTase; 1.
CC
CC KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC
CC FT TRANSMEM 1 21
CC FT DOMAIN 22 435
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SEQUENCE 435 AA; 49976 MW; DE26E49DEBEE629 CRC64;
CC
CC Query Match 43.3%; Score 1302; DB 1; Length 435;
CC Best Local Similarity 51.0%; Pred. No. 2.9e-97;
CC Matches 268; Conservative 57; Mismatches 97; Indels 104; Gaps 9;

```



```

OY 391 LLEPDLVFLVIR-----SRHVNPTRKRAVLVYSLNMGYAENLKSMTWEPTSTGELIIGHOPS 447
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 350 LLEPDLVIR--ORNLPSGOVLNRKSKRAVFISSSSPGCYEISIDVYMENTVNGEIIIVHKPS 407
OY 448 QECYQOTERKMHNGKLAEMAYLLSTLDNL-VTSAMSTFGYVAGGLGKPMIILYRPENRT 506
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 408 YKQYOTPRMESEKRAMAEIYLLSCDALVYTGMLSSLVAVAGLGLKPMYLNKENG 467
OY 507 TPDPSGGRMSMEPCFHSPPFYDCK 531
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 468 AHEPYCVKARSLIEPCSOATLEHGCK 492

RESULT 12
NODZ_RHISN
ID NODZ_RHISN STANDARD: PRT: 322 AA.
AC P55355;
AT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Modulation protein 2 (EC 2.4.1.-).
GN NODZ OR Y4AH.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; Pubmed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -I- FUNCTION: FUCOSYLTRANSFERASE WHICH ADDS THE FUCOSE MOIETY OF THE
CC GDP-FACTOR ON ITS TERMINAL REDUCING N-ACETYLGLUCOSAMINE END. USES
CC GDP-FUCOSE AS THE DONOR GROUP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE000064; AAB91605.1; -
KW Modulation; Transferase; Glycosyltransferase; Plasmid.
SQ SEQUENCE 322 AA; 36523 MW; 07800CF24CFCFB7B CRC64;

Query Match 3.7%; Score 110.5; DB 1; Length 322;
Best Local Similarity 20.9%; Pred. No. 0.17;
Matches 74; Conservative 50; Mismatches 123; Indels 107; Gaps 16;

OY 172 KYVWVWISFSGNRIISLASYFLYALLTDRLVLD-RGKMDMDLFCFEPFLGMSWLLPLDF 230
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 4 RYVLSRRRTGGDCILMSLAAMRYAQRFLAVDMRGS-----CY-----LDQ 47
OY 231 PMTQGFQSLNDESSRCYCYMYKNOVYIDEGTSLHYLHLVHDYGDHDMKFCF---GDQT 287
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 48 PFTNAFP-----VFEPPIKDIAGV-----PFTCDNRVNEFS 78
OY 288 FIGVPMVLIVKTDNVFVPSLILIGFD-----DELNKLFPQKATV----- 327
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 79 FPGP-----FPFMWNNKRAIECVYPPDAQVFRERBELDELFOADQDVEANTVYC 127
OY 328 -----FHHLGRLV---HPTNQGVLVTRYEYAVLSHADEKIGIQVRFDEDPGFQ 376
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 128 DACLMMRCDEAEAKQIFCSVPRAEIOARIDAIYQEHF-YGSAIGVHVR-----HGNGE 181
OY 377 HVMDQISSCQKEXKLEPVDLVERSRRHVNPTRKRAVLVYSLNMGYAENLKSMTWEPTS 436
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 182 DVMDBATVMDPDLAIVAOVCTAIINAARALPRPKFVRYVLTGDSARVLVDQVSSRPDLTL 240

```

```

OY      437 TGEITGVHOSQEG-YOOTERKMHNG-KALAEVLLSLTNIV-----TSMSTP 484
Db      241 ---IPKFRADQSGLHSADLGVGCGISALVENYLLGLCTVTIRPPTSAFTKY 291

RESULT 13
ID       UVRC_HELPJT          STANDARD;           PRT;         594 AA.
AC       G9ZL21;
DT       30-MAY-2000 (Rel. 39, Created)
DT       30-MAY-2000 (Rel. 39, Last sequence update)
DT       15-JUN-2002 (Rel. 41, Last annotation update)
DE       Excnuclease ABC subunit C.
GN       UVRC OR JHP0760.
OS       Helicobacter pylori j99 (Campylobacter pylori j99).
OC       Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX       Helicobacter.
NCBI_TaxID=85963;
RN       [1]
RP       SEQUENCE FROM N.A.
RX       MEDLINE=99120557; PubMed=9923682;
RA       Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Dolg P.C.,
RA       Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA       Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.W., Ives C.,
RA       Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RT       Trust T.J.;
RT       "Genomic sequence comparison of two unrelated isolates of the human
RT       gastric pathogen Helicobacter pylori.";
RL       Nature 397:176-180(1999).
CC       -I- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
CC       UVRA-UVRB COMPLEX, REPLACING UVRA, AND THE DAMAGED DNA STRAND IS
CC       NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).
CC       -I- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC       -I- SUPRACELLULAR LOCATION: CYTOPLASMIC (By similarity).
CC       -I- SIMILARITY: BELONGS TO THE UVRC FAMILY.
CC       -I- SIMILARITY: CONTAINS 1 UVR DOMAIN.
CC       -----
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC       or send an email to license@lsb.slb.ch).
CC       -----
DR       EMBL, AE001506; AAD06337.1; -.
DR       InterPro: IPR001943; UVRB/C.
DR       InterPro: IPR004791; UVRC.
DR       InterPro: IPR001162; UVRC_C.
DR       InterPro: IPR003305; UVRC_N.
DR       Pfam: PF01541; Excl_endo_N_1.
DR       ProDom: PD005870; UVRC_C; 1.
DR       SMART: SM00465; GYC; 1.
DR       TIGRFAMS: TIGR00194; uvrc; 1.
DR       PROSITE: PS50151; UVR; 1.
KW       SOS response; Excision nuclease; DNA repair; Complete proteome.
FT       DOMAIN             205      240      UVR
SEQUENCE   594 AA; 68671 MW; 0B63C45BA42342E7 CRC64;

Query Match              3.2%; Score 97; DB 1; Length 594;
Best Local Similarity    20.8%; Pred. No. 4.9;
Matches     55; Conservative 39; Mismatches 103; Indels 68; Gaps 11;
```

```

      RESULT 13
UVRC_HELPJT ID STANDARD; PRT; 594 AA.
AC Q9ZL21;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Excnuclease ABC subunit C.
GN UVRC OR JHP0760.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
CC Helicobacter.
CX NCBJ_TaxID=85963;
RN [1]
RX SEQUENCE FROM N.A. MEDLINE=99120557; PubMed=9923682;
RA Smith R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Altm D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Metberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -! FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC CARLYZES THE EXCISION REACTION OF UV-DAMAGED NUCEOTIDE SEGMENTS
CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
CC UVRA-VVRB COMPLEX, DISPLACING VVRA, AND THE DAMAGED DNA STRAND IS
CC NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).
CC -! SUBUNIT: CONSISTS OF THREE SUBUNITS: UVRA, VVRB AND UVRC.
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC (By similarity).
CC -! SIMILARITY: BELONGS TO THE UVABC FAMILY.
CC -! SIMILARITY: CONTAINS 1 VVR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isd-sdb.ch).
CC -----
DR EMBL; AE001506; AAD06337.1; -
DR InterPro: IPR001943; VVRB/C.
DR InterPro: IPR004791; UVRC.
DR InterPro: IPR001162; UVRC_C.
DR InterPro: IPR000305; UVRC_N.
DR Pfam: PF01541; Excl_endo_N_1.
DR Pfam: PF02151; VVR; 1.
DR ProDom: PD003870; UVRC-C; 1.
DR SMART: SM00465; GYC; 1.
DR TIGRFAMS: TIGR00194; uvrc; 1.
DR PROSITE: PS01511; VVR; 1.
KW SOS response; Excision nuclease; DNA repair; Complete proteome.
FT DOMAIN 205 240 VVR.
FT SEQUENCE 594 AA; 68671 MW; 0B63C45BA42342E7 CRC64;
Query Match 3.2%; Score 97; DB 1; Length 594;
Best Local Similarity 20.8%; Pred. No. 4.9;
Matches 55; Conservative 39; Mismatches 103; Indels 68; Gaps 11;
OY 78 ARVDAGFPWNTNINSKLL-----GGLAS-----GEDESCLSRYSQVHYRK 122
|:::|:| | | :| | | | | | | |
DB 249 AKLVLDLFATFYGGNNKAVLKVMFMRGKKIITSSAFETIHSLNGFEPTDAMKOAIINHQS 308
|:::|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
123 PSPKPSPSYLSKLKNYEKLHKGCGPGETSEYKKALKOLDOEHIOGDGCCKVVMISFGSL 182

```


Db 309 HLPMPQILLSCSN-----ETLKL-OEFLSHQYSKIALSIPKGD 351
 QY 183 GNRILSLA-----SVFLYALLTDRVLLVDRGKMDLDFCEPFLGMSWLLPLDFPMTDOF 237
 Db 352 KLALIEIAMNAQEIFSQEKTNSNEDRLERSLFJNECPY-----RVLEIFD 399
 QY 238 GLNOESRCYGVW--KNQYIDTEGLSHLYLHVHDYGDHDKM--FFCEGDQTFICKVP 293
 Db 400 TSHHSNSQCVGQVWYENMFQKD---SYRRYHL-KGSNEYDQMSELLTRALDFAKEPP 455
 QY 294 WLIVKTDNYFVPSLWLLRFGFDELIN 318
 Db 456 -----PMLWIDGGRQILN 469

RESULT 14

MDRB_HUMAN STANDARD; PRT; 1224 AA.

AC 09BZH6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE WD-repeat protein 11.
 GN WDR11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21427518; PubMed=11536051;
 RA Chernova O.B., Hunyadi A., Malaj E., Pan H., Crooks C., Roe B.,
 RA Cowell J.K.;
 RT "A novel member of the WD-repeat gene family, WDR11, maps to the 10q26
 RT region and is disrupted by a chromosome translocation in human
 RT glioblastoma cells.";
 RL Oncogene 20:5378-5392(2001).
 CC -1- TISSUE SPECIFICITY: Ubiquitous.
 CC -1- DISEASE: A form of glioblastoma is characterized by a chromosomal
 CC translocation t(10;19)(q26;q13.3) which involves WDR11 and ZNF320.
 CC -1- SIMILARITY: CONTAINS 9 WD REPEATS (TRP-ASP DOMAINS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF320223; AAK08064.1; -;
 DR MIM: 606417; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF004400; WD40; 4.
 DR SMART: SM00320; WD40; 6.
 DR PROSITE: PS00678; WD_REPEATS_1; 3.
 DR PROSITE: PS50082; WD_REPEATS_2; FALSE_NEG.
 DR PROSITE: PS50294; WD_REPEATS_REGION; FALSE_NEG.
 KW Repeat; WD repeat; Transmembrane; Proto-oncogene;
 KM Chromosomal translocation.
 FT REPEAT 59 108 WD 1.
 FT REPEAT 111 154 WD 2.
 FT REPEAT 354 393 WD 3.
 FT REPEAT 471 510 WD 4.
 FT REPEAT 566 605 WD 5.
 FT REPEAT 708 745 WD 6.
 FT REPEAT 747 787 WD 7.
 FT REPEAT 793 831 WD 8.
 FT REPEAT 893 940 WD 9.
 FT TRANSMEM 1127 1147 POTENTIAL.
 SQ SEQUENCE 1224 AA; 136683 MW; 9182221ABAEFBA4ED CRC64;

Query Match

3.2%; Score 97; DB 1; Length 1224;

Best Local Similarity 18.9%; Pred. No. 13;
 Matches 79; Conservative 61; Mismatches 149; Indels 128; Gaps 18;

QY 152 SYKKALKQLODHEIDHGGGECKYVWISPSGKGNLLSLASFVLLALLDRVLLVDRCKDM 211
 Db 824 SMKACFRMDQDELTEPYWCPLYL-----LVPRASIALKNAFLLH----- 861
 QY 212 DDLCEPFLGMSWL--PLDFPMTDOFGLNOESSRCYGVWKNQVDTGTLTSLYHL 269
 Db 862 -----QPMNGOYSIDSHVDYPRNEMELKNLQBDNLNSIDIKKLLDPEFTLQROCLV 916
 QY 270 VHDYGDHDKMFCEGDQTFICKVPWLIVKTDNYFVPSLWLLRFGFDELINLKEPQATVFH 329
 Db 917 SRLYGDSESLHF-----WTVA--HYLHSL-----SQEKASSTAPKREA--- 954
 QY 330 HLGRYLFRPTNQVWGLVTRYVEATLSHAD-----EKIGID-VAVFDEPPGPRHVNDOI 382
 Db 955 -----PRDKLSNPIDICVDLCENAVYFQKFLQERVNLQEVK-----RSTYDHT 997
 QY 383 SSCYCKEKLPEVDLVE-----RSRHVNTPKKAVLWLSNAGYAENLKSMEWEPY 435
 Db 998 RKCTDQILLIGQTPRAVQLLLETSDADNQHVCDSLKACLVTV-----T 1041
 QY 436 STGEIIGVHQPQSGGYQTEKKM-HNGKALAEWYLLSTDLNLTSA-----WSTFGY 486
 Db 1042 SSG-----PSQSTIKLVATNMIAKGLACGVOLLCIDKAADACRYLQTYGEMNRAM 1094
 QY 487 VA-----QGLGGLKMWI--LYPENRTTDP-----SCGRAMSMEPCFHSPPFYD 529
 Db 1095 LAKVLNPEECADVLRRWVDHLCSPQVOKSKALLVLISLCFSPSVAETLHSMRYFD 1151

RESULT 15

UVR_C_HELPY STANDARD; PRT; 594 AA.

AC P56428;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Exonuclease ABC subunit C.
 GN UVR_C OR HP0821.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirknes E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: THE ABC EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
 CC UVR_A-UVR_B COMPLEX, REPLACING THE UVR_A, AND THE DAMAGED DNA STRAND IS
 CC NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVR_A, UVR_B AND UVR_C.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE UVR_C FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 UVR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis/ebis-licensing/> or send an email to license@ebi.ac.uk).

DR EMEL; AE000593; AAD07868.1; -,
DR TIGR; HP0821; -,
DR InterPro; IPR001943; VYTB/C,
DR InterPro; IPR004791; VYTC,
DR InterPro; IPR001162; VYTC_C,
DR InterPro; IPR000305; VYTC_N,
DR Pfam; PF01541; Excl_endo_N_1,
DR Pfam; PF02151; VYR; 1,
DR ProDom; PD005870; VYTC_C; 1,
DR SMART; SM00465; GYTC; 1,
DR TIGRFAMS; TIGR00194; VYTC; 1,
DR PROSITE; PS0151; VYR; 1,
DR SOS response; Excision nuclease; DNA repair; Complete proteome
FT DOMAIN 205 240 VYR.
SQ SEQUENCE 594 AA; 68679 MW; 2550B3FE45FBAC6 CRC64;

Query Match	3.2%;	Score 96;	DB 1;	Length 594;
Best Local Similarity	20.0%;	Pred. No. 5.9;		
Matches	53;	Conservative	41;	Mismatches 103;
				Indels 68;
				Gaps 10;

Oy	78	ARVLDAGEFPNNTINISDKLL-----GGLLAS-----GDDESCSLRVSQVHYRK	1222
Dd	249	AKLIDLDIFATRYGASNKAVLYLMKMRGGKITISSAFEKIHSHINGPDTBANKOALIIINYOS	3088
Oy	123	PSPYKPSSYYLSIKLTNNYEKLHRCRGPGTSEYKALKOLDOEHINDGEDCKVYVAISFSGI	1828
Dd	309	HLPMPGEOILLMACSN-----ETLKEL-DEFISHQSKIALSIPKGD	3521
Oy	183	GNRILSLA-----SYELVALTLDRVLYVDRCDDDLFCPEFLGMSWLLPLDPMTDQED	2327
Dd	352	KLALETIMKMAOELEFSOEKTSDNLILLEARSJLFKLECMPLY-----RYEITFD	3599
Oy	238	GLNOESSRCYGVYMNKNQVIDTEGFLSHLYLVLYHDYGDHDKMFECEGDOTFIGKVPALIV	2927
Dd	400	TSHSSSQCGVMV---VEENNAFOKNSY-----RRHLKGSDEYT-QOMSELLT	4444
Oy	298	KTDNYEV---PSLMLDPGFDELIN	318
Dd	445	RRAALDFAKEPPNMLVINDIGGRACLN	469

Search completed: February 27, 2003, 15:31:00
Job time : 17 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Computex Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 10:24:25 ; Search time 4332 Seconds

(without alignments)
1165.481 Million cell updates/sec

Title: US-10-037-311a-2

Perfect score: 1662

Sequence: 1 atgatacgaattcgtacag.....ggggacttaagctatgta 1662

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: GenEmbl.*
2: gb_da.*
3: gb_htg.*
4: gb_in.*
5: gb_om.*
6: gb_ov.*
7: gb_pat.*
8: gb_ph.*
9: gb_pl.*
10: gb_pr.*
11: gb_ro.*
12: gb_sts.*
13: gb_sy.*
14: gb_un.*
15: gb_vl.*
16: em_ba.*
17: em_fun.*
18: em_hum.*
19: em_in.*
20: em_mu.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pl.*
27: em_ro.*
28: em_sts.*
29: em_un.*
30: em_vl.*
31: em_htg_hum.*
32: em_htg_inv.*
33: em_htg_other.*
34: em_htg_mus.*
35: em_htg_pln.*
36: em_htg_rnd.*
37: em_htg_mam.*
38: em_htg_vrt.*
39: em_sy.*
40: em_htgo_hum.*
41: em_htgo_mus.*
41: em_htgo_other.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1632.2	98.2	1768	8	AF154111 Arabidops
2	1401.6	84.3	91318	8	AC005313 Arabidops
3	695.2	41.8	99547	8	AC068197 Arabidops
4	695.2	41.8	112126	8	AC007576 Arabidops
5	642.6	38.7	1698	8	AF223643 Pisum sat
6	607.4	36.5	89301	8	AC006920 Arabidops
7	606	36.5	1512	6	AX412254 Sequence
8	606	36.5	1512	6	AX412452 Sequence
9	606	36.5	1512	6	AX412825 Sequence
10	606	36.5	1512	6	AX417474 Arabidops
11	604	36.3	1602	8	AF417475 Arabidops
12	554.6	33.4	160096	2	AP004817 Oryza sat
13	373.6	22.5	147289	2	AP005002 Oryza sat
14	344.4	20.7	134402	8	AC011765 Arabidops
15	344.4	20.7	1706	8	AF417473 Arabidops
16	336.2	20.2	162342	2	AP004687 Oryza sat
17	336.2	20.2	187916	2	AP003458 Oryza sat
18	336	20.2	162342	2	AP004687 Oryza sat
19	336	20.2	187916	2	AP003458 Oryza sat
20	229.8	13.8	82771	2	AP003418 Oryza sat
21	150	9.0	147227	8	OSJN00013 Arabidops
22	94	5.7	149120	2	AP005536 Oryza sat
23	93.2	5.6	151133	2	AP004770 Oryza sat
24	93.2	5.6	151133	2	AP004770 Oryza sat
25	93.2	5.6	188228	2	AP004771 Oryza sat
26	71.8	4.3	148669	2	AC091532 Oryza sat
27	55	3.3	150116	2	AP004816 Oryza sat
28	55	3.3	153717	2	AP004841 Oryza sat
29	49	2.9	151163	8	AC079128 Oryza sat
30	48.4	2.9	1141	6	AX083744 Sequence
31	45.4	2.7	144563	9	AC006204 Homo sapi
32	43.6	2.6	7218	6	166494 Sequence 14
33	43.4	2.6	7001	6	AX347354 Sequence
34	43.4	2.6	7001	6	AX349075 Sequence
35	42.4	2.6	176486	8	AP003768 Oryza sat
36	42.4	2.5	186305	9	AL356098 Human DNA
37	41.4	2.5	3625	3	AF081801 Dictyoste
38	41.2	2.5	91001	8	AC013453 Arabidops
39	41.2	2.5	159581	2	AC108667 Homo sapi
40	41.2	2.5	170285	2	AC069237 Homo sapi
41	41	2.5	19395	3	CER0381 Caenorhabdi
42	41	2.5	87081	8	AB073164 Arabidops
43	41	2.5	112684	8	AP002033 Arabidops
44	41	2.5	165458	2	AC009795 Homo sapi
45	41	2.5	171962	9	AC091824 Homo sapi

ALIGNMENTS

```
RESULT 1
AF154111 1768 bp mRNA linear PLN 28-JUN-1999
LOCUS Arabidopsis thaliana xyloglucan fucosyltransferase (FTI) mRNA,
DEFINITION complete cds.
ACCESSION AF154111
VERSION AF154111.1 GI:5231144
KEYWORDS SOURCE
ORGANISM Arabidopsis thaliana.
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Perrin,R.M., DeRoche,A.E., Bar-Peled,M., Zeng,W., Norambuena,L.,
```

Orrellana, A., Raikhel, N.V. and Keegstra, K.
Xyloglucan fucosyltransferase, an enzyme involved in plant cell
wall biosynthesis
Science 284 (5422), 1976-1979 (1999) *True*
JOURNAL 99301928
MEDLINE 10373113
PUBMED 2 (bases 1 to 1768)
REFERENCE Perrin, R.M., DeRoche, A.E., Bar-Peled, M., Raikhel, N.V. and
AUTHORS Keegstra, K.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-1999) MSU-DOE Plant Research Laboratory, Michigan
State University, Room 110 Plant Biology Building, East Lansing, MI
48824, USA
FEATURES
source Location/Qualifiers
1..1768
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="BAC T18E12; C1C11A04"
/clone="EST 191A67"
/clone_lib="lambda ZapII (Dr. J. Kleber, Dr. J. Ecker and
associates)"
/dev_stage="3 day old seedling"
1..1768
/gene="Frl1"
19..1695
/function="transfers Fuc to Gal in alpha-(1,2) linkage on
the plant cell wall polysaccharide xyloglucan"
/product="xyloglucan fucosyltransferase"
/protein_id="AAD41092.1"
/db_xref="GI:5231145"
/translation="MDNSYRRSSRPITTTGGSGSYNSELLOKIIISGTMKTRT
FTTCLIVSVIVAFSMIRHQRHSDSNRIRNGEAEAVLDAGVFPNVTINISDKLGL
ASGRPDSLSIRYQSVHTRKPSYPKPSYLSIKLNRKRCGPSTESYKALKOL
DOEHIDGDECKVWVWISFSGNRIELASVFLALITDRLVDRGDMDDLECEP
FLGMSMLPLDPMPTDQFDGLNDESSRCYGVKNQVIDETGSLYLHFDHGDH
DKMFCEDQPTFIRYVPLIVTDNYFVPSLMLDPGDEDELNKLEPOKATVFNHNGRY
LEHPNQWGLVTRYEAVLYSHADEKIGIORYVDEDPGPRHWDQISCTQCEKL
FEVDLVERSHVNTPKRAVLYTSLMGVAMENLKSMTWEYPTSGELIGVHPSQBS
YQOETKMHNKALAEWYLSTDNLYTSANSTFVYAGLGGLKPMVLTVEPNTTP
DPSGGRASMECPSPFPIDCKAKTGIDTGLTVPHVHNCEDISWGLKLV"
BASE COUNT 487 a 359 c 378 g 544 t
ORIGIN
Query Match 98.2%; Score 1632.2; DB 8; Length 1768;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1659; Conservative 0; Mismatches 3; Indels 15; Gaps 1;
QY 1 ATGGATCAGATTGCTACAGAGAAAGTCGTCGATCAGAACCACTACGCGGCTTCA 60
DB 19 ATGGATCAGAAATTCGTACAGAGAAAGTCGTCGATCAGAACCACTACGCGGCTTCA 78
QY 61 AAGTCGGTAATTCGCGAACTACTCAATGAAGATCTAGCTCCGGATCGATGAAG 120
DB 79 AAGTCGGTAATTCGCGAACTACTCAATGAAGATCTAGCTCCGGATCGATGAAG 138
QY 121 CTCACGAGAACCTTCAGTCTGATAGTCTCTGATAGTATGATGCAATTCGATG 180
DB 139 CTCACGAGAACCTTCAGTCTGATAGTCTCTGATAGTATGATGCAATTCGATG 198
QY 181 ATCTTTACCAACACCCATCTGATTCGAATGGATTATGGTTTCGCGAAGCTAGAGT 240
DB 199 ATCTTTACCAACACCCATCTGATTCGAATGGATTATGGTTTCGCGAAGCTAGAGT 258
QY 241 CTCGACGCCGAGTTTCCCA-----AATTGATGAAGCTTCTGGAGGG 285
DB 259 CTCGACGCCGAGTTTCCCAAAATGTTACTAACAATTCGATGAAGCTTCTGGAGGG 318
QY 286 CTACTGCTTCTGTTTGGATGAAGATTCCTGCTAGTAGTACCAATCGTTCATATAC 345
|||||

DB 319 CTACTGCTTCTGTTTGGATGAAGATTCCTGCTAGTAGGACCAATCACTTCATTAC 378
QY 346 CGTAAACCTTCACCTTACCAAGCAATCTTCTATCTCATCTTAAGCTTGAAGATCAGAA 405
DB 379 GTTAAACCTTCACCTTACCAAGCAATCTTCTATCTCATCTTAAGCTTGAAGATCAGAA 438
QY 406 AAGCTTCACCAAGCAATGTCGCGGCTAGTCAATCTTACCAAGCAATCTTGAAGCTT 465
DB 439 AAGCTTCACCAAGCAATGTCGCGGCTAGTCAATCTTACCAAGCAATCTTGAAGCTT 498
QY 466 GATCAGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 525
DB 499 GATCAGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 558
QY 526 GGTATGAGGAGACGATCTTCTAGCCGCTGCTTCTAGCCGCTTCTAGCCGCTTCTAGCC 585
DB 559 GGTATGAGGAGACGATCTTCTAGCCGCTGCTTCTAGCCGCTTCTAGCCGCTTCTAGCC 618
QY 586 AGAGCTTCTGCTTGTGACCGAGGAAAGACATGATGATGATGATGATGATGATGATG 645
DB 619 AGAGCTTCTGCTTGTGACCGAGGAAAGACATGATGATGATGATGATGATGATGATG 678
QY 646 GGTATGCTGCTGCTTGTGACCGAGGAAAGACATGATGATGATGATGATGATGATGAT 705
DB 679 GGTATGCTGCTGCTTGTGACCGAGGAAAGACATGATGATGATGATGATGATGATGAT 738
QY 706 CAGAATCATCTCTGCTTATGATATGATGATGATGATGATGATGATGATGATGATGAT 765
DB 739 CAGAATCATCTCTGCTTATGATATGATGATGATGATGATGATGATGATGATGATGAT 798
QY 766 ACTTGTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 825
DB 799 ACTTGTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 858
QY 826 TTCTGTGAGAGAGACCAACATTCATGCGGAGAACCTTCTGCTGATTTGTTAAACGAC 885
DB 859 TTCTGTGAGAGAGACCAACATTCATGCGGAGAACCTTCTGCTGATTTGTTAAACGAC 918
QY 886 AATTACTTGTTCATCTCTGCTGTTAATATACCGGCTTTCATGATGATGATGATGATG 945
DB 919 AATTACTTGTTCATCTCTGCTGTTAATATACCGGCTTTCATGATGATGATGATGATG 978
QY 946 TTCCACAGAAAGGACGCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1005
DB 979 TTCCACAGAAAGGACGCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1038
QY 1006 CAAATGAGGCTTACTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1065
DB 1039 CAAATGAGGCTTACTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
QY 1066 ATTGGATTCAGTAAGATTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1125
DB 1099 ATTGGATTCAGTAAGATTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1158
QY 1126 CAGATTCATCTTGTACTCAAAAGAGAACTTCTACCTGAATATACACACACATGAGAG 1185
DB 1159 CAGATTCATCTTGTACTCAAAAGAGAACTTCTACCTGAATATACACACACATGAGAG 1218
QY 1186 AGATCTCGCATGTTAATATACCCCAACACAAAGCCGCTGCTTTCATCTTGAAGCGG 1245
DB 1219 AGATCTCGCATGTTAATATACCCCAACACAAAGCCGCTGCTTTCATCTTGAAGCGG 1278
QY 1246 GGTATGAGGAGAACTTAAAGATATGATGATGATGATGATGATGATGATGATGATGAT 1305
DB 1279 GGTATGAGGAGAACTTAAAGATATGATGATGATGATGATGATGATGATGATGATGAT 1338
QY 1306 ATCGTGTTCATCAGCGCAGCAAGAGATTTATCAGCAGACCGAAAGAAAGATGATAT 1365
DB 1339 ATCGTGTTCATCAGCGCAGCAAGAGATTTATCAGCAGACCGAAAGAAAGATGATAT 1398
QY 1366 GGCAGAGCTTCTCGGAATGATCTTGTGATGATGATGATGATGATGATGATGATGAT 1425
DB 1399 GGCAGAGCTTCTCGGAATGATCTTGTGATGATGATGATGATGATGATGATGATGAT 1458
|||||

```
OY 1426 TGGTCTACATTGGATGATGATGAGCTCAGGCTCTGGAGGTTTAAAGCCTTGACTCTAT 1485
|||||
DB 1459 TGGTCTACATTGGATGATGATGAGCTCAGGCTCTGGAGGTTTAAAGCCTTGACTCTAT 1518
|||||
OY 1486 AGACCCGAAAAACCGTACACACTCCGATCCCTGCTGCTGGGCTATGTCGATGAGCCT 1545
|||||
DB 1519 AGACCCGAAAAACCGTACACACTCCGATCCCTGCTGCTGGGCTATGTCGATGAGCCT 1578
|||||
OY 1546 TGGTTCGACTGGCCTTCATTCATGATGTAAGGAAAAACGGGATTTGACACGGGAACA 1605
|||||
DB 1579 TGGTTCGACTGGCCTTCATTCATGATGTAAGGAAAAACGGGATTTGACACGGGAACA 1638
|||||
OY 1606 CTAGTTCCTCATGTGAGACATTTGTGAGCATATCAGCTGGGACCTTAAGCTAGTATGA 1662
|||||
DB 1639 CTAGTTCCTCATGTGAGACATTTGTGAGCATATCAGCTGGGACCTTAAGCTAGTATGA 1695
|||||

RESULT 2
AC005313 91318 bp DNA linear PLN 11-MAR-2002
LOCUS complete sequence.
DEFINITION Arabidopsis thaliana chromosome 2 clone T18E12 map C1C11A04,
ACCESSION AC005313
VERSION AC005313.3 GI:20197340
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 91318)
AUTHORS Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
Shen,M., Rominig,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 91318)
AUTHORS Lin,X.
JOURNAL Direct Submission
TITLE Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 91318)
AUTHORS Town,C.D. and Kaul,S.
JOURNAL Direct Submission
TITLE Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr., Rockville, MD 20850, USA, cdtown@igr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6558460.
FEATURES
SOURCE Location/Qualifiers
1..91318
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C11A04"
/clone="T18E12"
/complement(2646..2678)
/rpt_family="(A)n"
/repeat_region complement(2800..2854)
/repeat_region complement(3183..3217)
/repeat_region complement(3183..3217)
/rpt_family="AT-rich"
/complement(5635..6198)
/gene="At2g03320"
/feature="T18E12.5; predicted by genscan"
/complement(join(<5635..5876,6135..>6198))
/gene="At2g03320"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA015037.1"
/db_xref="GI:20197347"
/translating="MLINK:KIN7YHYVFLCMASKVSEKLSFRKAERPNLPJIA
ATTLISPOGRVDFVMDNEAOEIRNVSPISOELPSSSTTSSHSNRSNRSIVIO"

repeat_region complement(6171..6195)
rpt_family="AT-rich"
complement(6248..6271)
rpt_family="(TA)n"
repeat_region complement(6305..6371)
rpt_family="AT-rich"
gene 6491..6949
/gene="At2g03310"
/feature="T18E12.6; predicted by genscan"
<6491..>6949
/gene="At2g03310"
6491..6949
/gene="At2g03310"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA015038.1"
/db_xref="GI:20197348"
/translating="MNRBAEAEQEEFSEWVVIQTSKSPSPRTDASPNSPILOPSNR
RENPTIINDDSDSVSPSTVNLMPRVLETTKRLIKNSVCIFQAVRVRCSYLRKR
VFMSLTIIGFSLVLSLVYVKLVMMWRRLHEKLRFLLLLRKDYCVN"
8920..8985
/rpt_family="AT-rich"
9006..9072
/rpt_family="AT-rich"
9167..9187
/rpt_family="AT-rich"
9358..10112
/gene="At2g03300"
/feature="T18E12.2; domain signature TIR exists, suggestive
of a disease resistance protein"
join(<9358..9563,9631..9876,9953..>10112)
/gene="At2g03300"
join(9358..9563,9631..9876,9953..10112)
/gene="At2g03300"
/codon_start=1
/product="disease resistance protein (TIR class),
putative"
/protein_id="AA015034.1"
/db_xref="GI:20197344"
/translating="MTFSSPTQVFLNWRGEOLRSPVSHLIDAFERNEINFVUDKYEO
RGKDLNLFKRIQESKIALALESTRYTESWCMDLVKIKKLADKRLHVIPIYKVK
VEDVRKQGTGEFNDWFLAKVSGDOIKKWEALCEIIPNKGSLIGDKSDEADFIKEY
VKAVOCVATATGLEEENHFGKKRKDCCKEPLDLKRSRTKTL"
complement(11109..11139)
/rpt_family="AT-rich"
complement(11148..11178)
/rpt_family="(TA)n"
complement(11205..11293)
/rpt_family="AT-rich"
complement(113046..14158)
/gene="At2g03290"
/feature="T18E12.3"
complement(join(<13046..13290,13406..13482,13662..13871,
13949..>14158))
/gene="At2g03290"
complement(join(13146..13290,13406..13482,13662..13871,
13949..14158))
/gene="At2g03290"
/codon_start=1
/product="putative Golgi-associated membrane trafficking
protein"
/protein_id="AA015035.1"
/db_xref="GI:20197345"
/translating="MDICRSSILLLIALLSPRTLMSRYELKSKTKICGEETHENAM
SICKYFTVNPENEDHPLPDSHKITIVKVPDQKNLHBDKVEAGQFSFTAYENGSSVYA
CITADIVKPEPTTLIDFDWKGTGVSKEMTVNAKSOYDMMEYQVYTLADTVISHEEM
YYLREEREEMOELNRSTNSKMAWSPFSGVLVCLSVAGLQFWHLKTFPEKKKLI"
complement(14743..14793)
/rpt_family="AT-rich"
15341..17331
/gene="At2g03280"
/feature="T18E12.4"
join(<15341..15436,15491..15590,15680..15753,15817..15905,
```

```
15987..16149,16242..16531,16607..16811,16886..17043,
17109..>17331)
/gene="At2g03280"
join(15341..15436,15491..15590,15680..15753,15817..15905,
15987..16149,16242..16531,16607..16811,16886..17043,
17109..17331)
/codon.start=1
/gene="At2g03280"
/product="similar to axi 1 protein from Nicotiana tabacum"
/protein_id="AA15036.1"
/db_xref="GI:20197346"
/translation="MSVVGDRESSTLMRSDYKAPPSQAIKPAKRLQTFMHSQITGLT
NHISFSLPVQSVLPPLPPLPRNYSNGLLVSCGGLNQMAARVARDMVTARLNL
LTLYVELDKKSFADTSDPEDIDIKHIDSLRDVEYIIRLPKRYSKXGKFLFEP
PVSWSNDKXYLQOVLPPRSKRYIHFRSPSLANNGLSLDILRRCRVNPGILRET
PRIELGSKIVRILOOGRSFVALHRYMDMLASGCHGTGDEBELKMRATYPM
WREKIVSEERVOGICLPTPEAVLYKALGFQDDIYIAGETFGAARLADLKE
SPRIKVEMLDLPTELQOFONHSSQMALDIFVSASNTFIPYYGMMAVVGHR
YLGRKTLIDLRKRLVELLDLHNKTLISWDOFAVAVKADHGRRGEPTRHKVISVR
KEEDFYANPOECIS"
/complement(17450..19797)
/gene="At2g03270"
/notes="11812..1: supported by cDNA:
gi_15027926_gb_AY045820.1"
/complement(join(<17450..19527,19705..19797))
/gene="At2g03270"
/complement(17590..19509)
/gene="At2g03270"
/codon.start=1
/product="putative helicase"
/protein_id="AA15033.1"
/db_xref="GI:20197343"
/translation="MARKMSLEAFVSTMAPLIDMEKEAETISMTSGASRIETAOAK
GTTILNKCVQVOTGLMGRSLIEFOSNGDVLPAHKFGNHVVLINKSPDSSPLA
OGVYRLKDSSTIVVDEVPREGTLNLRLEKLANDEYRMRKTLIDLSKGTGCGA
SDIAPVLEGEKPSVSKDYKSPTPNKNDLSQKDAITKRLSKDVLFLHGPEGTGK
TTTYVELVLOEVKRGSKILACASNIANDIVEKLVPHKAVLVGHPARLIPVDLS
ALDAVKLKGDSGLANDIRKEMKLNKLLKAKRNRLRLOKRLFGKEERKQDL
AVSDVINKADVILITLGLALTRKLDNRFDVLIIDEAOALAEVCMWALKSGSCILA
GDHLDPPTIOSAEERKGLRTEFLERLADLYGDEIKSMILVOYRMHLEINMSKEL
YDNKLTASVSASHLEFDLENTKSSSEATLLVDPAGCMEMKKEEESYVNEGSA
EVAMAHAKRLMESVOPSDIGILIPYAAVWLRIILGKREKILKDMETSYDGOGE
KEATISMVRSNKKEVGLKDOARMVAATRSRQCCIVCDTETVSSDAFLKMIET
FEHGEYLASEYTN"
/21248..25474
/gene="At2g03260"
/notes="11812..7"
join(<21248..21521,21643..21984,22920..23311,23552..23681,
23760..23880,23973..24231,24320..24424,24497..24691,
24770..25066,25150..25286,25375..>25474)
/gene="At2g03260"
join(21248..21521,21643..21984,22920..23311,23552..23681,
23760..23880,23973..24231,24320..24424,24497..24691,
24770..25066,25150..25286,25375..25474)
/gene="At2g03260"
/codon.start=1
/product="unknown protein"
/protein_id="AA15032.1"
/db_xref="GI:20197341"
/translation="MKFGELSSOMVOEMOAVVNDYLTLLKEILKLEKTNPPPP
PNAHVPEGISRKMTLYRARGSLVOPRGKROSSRSEEDIDIEGRAPITYSST
HGLTTPILMAEBSGEVELYFPRIDDEFNRKEFKYKEDYKMDIKMLNKMDALIT
AFRYKVENPVGMWEKTVEMTRLASDIATSTALASTARTMRMPAOAMEMAIQ
EGSFRENEDEHDSRGATGVKTSLSNMRARAPIDVLDHIKNNKATPRSTI
KGYLNSSONEILFNRLNLEVEKEKLFALVEYOKRLKLSYFLVLFKILKYK
DKLMKLIORESTEIKHFANGHRKGNILRPMOKRREHNVTESTGSAFCISLIYA
LVAILRTKTRPEAHENTYMNTEPLSLFGFVLHTHTWVAIDIIYWKRYNVAATF
GCGGTGELGYROVLFGITIGTRALLCVGLNDLMEVAPKRNKRLPELILPLIAL
FVYLINPFRLLYSTRFTFLCHCLAAFLVLYLPDEFLOGLTISOVALSINFY
ICYIGWDFKRLKNTCEASEIYISLYIASVLYLSLLOCMRMTIEESLDGYNV
KLYLTVAISLRTAGYEVRNTKNPSTSHLVLAGSSSILAAVCTWDFHMDGLNK
TSKNRMLRDLILIPOKKYVFIAMILNVLFEMLOTILNEFEFLKQOTLAAVASLE
IMRGMNMFREVNEHNLNNYKFRAFKSVLP.LPENVDEDDKDD"
```

```
repeat_region complement(26234..26278)
Query Match 84.3%; Score 1401.6; DB 8; Length 91318;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 255 TTTCCCAATTCGTGATGATGCTTTCGGAGGCTCTTCGCTTCGTGTTTGTGAAGATTC 314
DB 41772 TTTTTCATATTCGTATGATGCTTCGGAGGCTCTTCGCTTCGTGTTTGTGAAGATTC 41831
QY 315 TTGCCTTAGTAGTACCAATCATGTTTCATTCACCTTAACCAAGCATCTTC 374
DB 41832 TTGCCTTAGTAGTACCAATCATGTTTCATTCACCTTAACCAAGCATCTTC 41891
QY 375 TTATTCATCTCTTAAGCTTTAACAACCTACGAAGACCTTCACAGCGATGTCGGGTAC 434
DB 41892 TTATTCATCTCTTAAGCTTTAACAACCTACGAAGACCTTCACAGCGATGTCGGGTAC 41951
QY 435 TGAATCTTACAGAAAGCTTAAACAACCTGATCAAGAACATTTGATGATGATGATG 494
DB 41952 TGAATCTTACAGAAAGCTTAAACAACCTGATCAAGAACATTTGATGATGATGATG 42011
QY 495 ATGCAATATGTGTGTGATTTCTTTTAGCGGCTTAGGAACAGATTAATTCCTAGC 554
DB 42012 ATGCAATATGTGTGTGATTTCTTTTAGCGGCTTAGGAACAGATTAATTCCTAGC 42071
QY 555 CTCGGTTTCTTACCGGCTTTAACGGATAGAGTCTTGCTTGTACCGAGGAAAGA 614
DB 42072 CTCGGTTTCTTACCGGCTTTAACGGATAGAGTCTTGCTTGTACCGAGGAAAGA 42131
QY 615 CATGAGATGATCTTTTTCGAGCGCTTTCGATGATGATGATGATGATGATGATGAT 674
DB 42132 CATGAGATGATCTTTTTCGAGCGCTTTCGATGATGATGATGATGATGATGATGAT 42191
QY 675 CCTATGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 734
DB 42192 CCTATGACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42251
QY 735 GGTAAAGATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 794
DB 42252 GGTAAAGATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42311
QY 795 TCATGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 854
DB 42312 TCATGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42371
QY 855 GAAAGTCCCTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 914
DB 42372 GAAAGTCCCTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42431
QY 915 ACCGGTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 974
DB 42432 ACCGGTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42491
QY 975 CTTAGTAGTATCTTTTACCCCACTAACCAAGTAGTGGGCTTAGTACTAGATAC 1034
DB 42492 CTTAGTAGTATCTTTTACCCCACTAACCAAGTAGTGGGCTTAGTACTAGATAC 42551
QY 1035 CGAAGCTTACTTATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1094
DB 42552 CGAAGCTTACTTATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42611
QY 1095 AGACCCGGGTCATTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1154
DB 42612 AGACCCGGGTCATTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 42671
QY 1155 ACTTCTACCTGAAGTACACACTAGTGAAGATGTCGCCATGTTAATACCCCAACA 1214
DB 42672 ACTTCTACCTGAAGTACACACTAGTGAAGATGTCGCCATGTTAATACCCCAACA 42731
QY 1215 CAAAGCGGTCTTGTACATCTTTGAACGGGGTTCACGGGAGAACTTAAAGATATG 1274
DB 42732 CAAAGCGGTCTTGTACATCTTTGAACGGGGTTCACGGGAGAACTTAAAGATATG 42791
```

QY 1275 TTGGGAATATCCGACATCACTGGAGAAATCATCGGTGTTTCATCAGCCGAGCAAGG 1334
|||||
Db 42792 TTGGGAATATCCGACATCACTGGAGAAATCATCGGTGTTTCATCAGCCGAGCAAGG 42851
|||||
QY 1335 TTATCAGACGACCCGAAAAAAGATGCATTAATGCGAAGCTCTTCCGGAATGATCTTTT 1394
|||||
Db 42852 TTATCAGACGACCCGAAAAAAGATGCATTAATGCGAAGCTCTTCCGGAATGATCTTTT 42911
|||||
QY 1395 GAGTTGACAGATATCTTGACAAAGTGTGTCATCTATTTGGATATGATGATCAAG 1454
|||||
Db 42912 GAGTTGACAGATATCTTGACAAAGTGTGTCATCTATTTGGATATGATGATCAAG 42971
|||||
QY 1455 TCTTGAGAGTTTAAAGCCTTGATCTATAGACCCGAAAAACCTACACCTCCGATCC 1514
|||||
Db 42972 TCTTGAGAGTTTAAAGCCTTGATCTATAGACCCGAAAAACCTACACCTCCGATCC 43031
|||||
QY 1515 TTCGTGTGTGGGCTATGTCATGATGAGACCTTGTTCACACCTGCTCCATCTATGATTTG 1574
|||||
Db 43032 TTCGTGTGTGGGCTATGTCATGATGAGACCTTGTTCACACCTGCTCCATCTATGATTTG 43091
|||||
QY 1575 TAAAGCGAAAAAGGATTTGACAGCGGAACACTGTCCTCATGATGACATTTGAGGA 1634
|||||
Db 43092 TAAAGCGAAAAAGGATTTGACAGCGGAACACTGTCCTCATGATGACATTTGAGGA 43151
|||||
QY 1635 TATCAGCTGGGACTTAACTAGTATGA 1662
|||||
Db 43152 TATCAGCTGGGACTTAACTAGTATGA 43179
|||||

RESULT 3
AC068197 99547 bp DNA linear PLN 28-JUN-2000
LOCUS AC068197
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F16A14 from
AC068197
VERSION AC068197.4 GI:7940274
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 99547)
Shun,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C., Chlou,J., Choi,E.,
Conn,L., Conway,A., Gonzales,A., Hansen,N., Howling,B., Koo,T.,
Lam,B., Lee,J., Lenz,C., Li,D., Liu,A., Liu,K., Liu,S.,
Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M., Yu,G.,
Federpiel,N.A., Theologis,A. and Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC F16A14 from
Chromosome I

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99547)
AUTHORS Ecker,J.R.
JOURNAL Direct Submission
TITLE Submitted (30-APR-2000) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 99547)
Ecker,J.R.
REFERENCE Direct Submission
TITLE Submitted (19-MAY-2000) Arabidopsis thaliana Genome Center,
JOURNAL Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 99547)
Chouk,R., Shun,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N.,
Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,D., Liu,A., Liu,J.,
Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaverl,A., Toriumi,M., Vaysberg,M.,

Yu,G., Davis,R., Federpiel,N., Theologis,A. and Ecker,J.
Direct Submission
JOURNAL Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On May 19, 2000 this sequence version replaced gi:7767679.
COMMENT
FEATURES
source
1..99547
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F16A14"
join(40..972,1061..1201,1297..1578)
/note="unknown protein; similar to Est gp1A197084.1"
/codon_start=1
/evidence=not_experimental
/product="F16A14.1"
/protein_id="AAF79391.1"
/translacion="MPGDRISELPESLISQILHLPLTKASVTSVLSRWKMLNLY
PGLDNCRDPEPNNNEKILDFIDRFLEQNESRLQKFKVDSRDKIIRSDRIGDA
ISGRIVLDESNTYRRADCTDTPCTEPFPLWLSKTLVSLKLSGLDEGPFY
LPCLKFMHLREWDSGTMNLEKLVSGCPVLEELIYLHDKLVLRVRSLSKFSI
PFRRLSLFRRVYQTFEIDAPGLVMSLAKADPFRIVYKNTLSLMDLDTKIVFG
WMPDPEDLPKRNELRDELTCISVRYHMLSHNTYKALDYSKVLKIFRNLSVEEA
FPSSLOPLPAFLSPFLKHLILETECPVPMKPELVNVPFCFVSTLEHVEIKGF
DMGRDMDKIASYFLSNSAVLKKLLSLFPGCQHSSESDIYELMLKLRSPRCIIIR
C"
join(2702..3920,4012..4193,4279..4347,4450..4743,
4956..5165,5257..5493)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F16A14.2"
/protein_id="AAF79392.1"
/db_xref="gi:8778384"
/translacion="MYSRRELEDELERYRSEMDGTRKVKISESLRCPFCYIDRRKRD
YOPDDLRLHAGISGSSRTKDRKARLALERMRKRLRPREPRSPGSDVSLPK
EEFTGKKRSTLTTEEGEFTTENSSPHYAKAPKPVSGDSRSGERLKEPSDKD
PFPSNEDKSYAKRPLCYLSAKKEDEPQRIGLSHGSAFTYTRQKLVSLCAGNDQY
VYHWRKGLANMRKTRFKRYAGESGSKIREDLKGFEPNHWVTPMLKODNYRSLTAL
AIVDFGEMEGFRNATMDFHFEVSQGRDHDLPBGDKLVGMVAKODNYRSLTAL
GDIHRLKQGLDKSVSGKEADQRTFTVLSNLNLTVRSNDLQMESIYKOTSSYLEK
RMKEKDEMINTHNEMKSIIMOOTADYLAISYEEHKAISOHLEAQRKEDEBNYLDLC
QAKKTRRKLQMKHKNKMLTOBONCADEMDRLAQQOQKDELKROVLEPEKID
AEOALELEEMKRDLYVMHMOGBEGEDSITKMITKELEKEDEMYEYOSLVO
TLVYKHGTINDELQDARKALIRSMRELTTTRYIGVKRNGALDETPFKVAKERYPAVE
ADKAKEELCSLMHELDSDAMWPIKVVYEDGIAKELENEDEKLOELRKELEGEVYAA
VTQALKEENENYSGRYIVPELMFNKONRKASIKGVYIVNSMKQKPKPKR"R"
complete(6320..8971)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F16A14.3"
/protein_id="AAF79418.1"
/db_xref="gi:8778410"
/translacion="MVRPRLPLSHVGLIRKOGSSRPVYVPALARTNLITSHSEYK
EGTDYKALELNDIGVLRVNSKDDPLAFLSKRLGNVTLRSVQATVATYRIVG
WGDKRLDTFLFELVRRGDEGRSVYDLAIKAIMESQVLLLRVSTALVAKVAND
MFDEAIDIFFRAYVSLGRAPDIKALNLSISMISGSDMVGWFEWTERIGLADAH
TYVIVVQALMNRNDKELEKLSRLITSETPNCPVFLNFTEGCLNMDIYAFLLQ
PLRANLILVQSDGLIYRKVYRGLCYEMREMDLAEVYLQMEKELIDPVSATIE
GHRKNMILPKRAVDVPMFKRRKRLNICYVSIIOCCQKMESEANDLPREPENT
SLDVCYNVAFDALGKGRKEAELELRMGTKIADPVINTTLTIGGCCLOGCSDA
FDLMTENDGTCTPDIVIYVNLAGLGTATNGLAQAEFLDKMNERGVKRLTYPPNMY
EGLIDAGELDKAEAFVSELEKSRSDNSQSCAAGFLDHAERETLFEPLPKVY
YFTLFTSLCAEAKETISKAQDLDRMMKIGVPEKSMYGLKIGACRVNRYAKAEFE
ILVTKTIVLPDEFYTIMNTYCRNLNPKQALAEEDMKRRVDRKDVVYVYSLNSDE
LDKMRKMDAPVDVYVYTMIRYCHINDLKKVYALFDMKRRRLVPOVYVYVYL
KNKPERNSREMKAFDVKPDVFTYVTLDMQCKIGDGEARITDIEGVDYDAB
YTLAIACCCKKRYEAKMIFDRMIEGVRDVPVYVYALLAGCCRNQFVLKAVLYE
MLEKGIPTKASLSAVYAKLAKAGLR"
join(9352..9383,9480..10419,10934..11089,11318..11412,

QY 916 CCGGTTTCATGATGAACATAAAGCTATTCCACAGAAAGCAGTCTCTTCATCAG 975
DB 62426 CCTAGTTCCAAACGAGCTATATCAAGCTATTCCCGCAGAAAGACCGGTCTTTCACCA 62485
QY 976 TTAGTAGCTATCTTTTTCACCACACTAACAGATATGGGCTTAGTCACTAGATATAC 1035
DB 62466 CTGGCTCGGTATCTTTTTCACCGCAGAAAGCTGGGGATATGGTACGACGATCTAC 62545
QY 1036 GAGCTTCTATTCATGATGCGGATGAGAAAGTGGATTCAGTAAAGTATTCATGAA 1095
DB 62546 AATGCTTATCTATCAAGACAGACGAGATACAGGATTCAGGATTCAGGATTCAGCAGA 62605
QY 1096 GACCGGGTCCATTCACCATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
DB 62606 CAACATAAATATTCACACAGCTATGATGATGATGATGATGATGATGATGATGATGAT 62665
QY 1156 CTTCACCTGAAGTACACACACTAGTG- - - - -AGAGATCTCGCATGTTAAATACCC 1209
DB 62666 CTGTACCTGATATTTGCTGCACAGAAAGACACAAATACATATCAATCTCTCC 62725
QY 1210 AAACCAAGCGGCTGTGACATCTTGAACGGGGTTACGGGAGACTTAAAGCT 1269
DB 62726 AAACCTTAAGCTGTTCGTGATACATCTGAAATCCAGATCTCAATTAACCTTAAAGAG 62785
QY 1270 ATGTATTTGGAAATATCCGACATCACTGAGAAATCATCGGTTCATCAGCCAGCCAA 1329
DB 62786 ATGTATTTGGAAACCCCAACTACAAAGACAGACATAGTAAGATTTATCAGCAGCAATGA 62845
QY 1330 GAAAGTTATCAGCAGACCCGAAAAAAGATGATTAATGSCAAAGCTCTTGGCGAAATGAT 1389
DB 62846 GAAAGTTTCAAGCAACAGACAAAGACCTTCAACGACCAAGGCGCTGACAGAGATGAC 62905
QY 1390 CTTTGAATTTGACAGATATCTTGACAGTGGTGTGATCTGATCTGATCTGATCTGATCT 1449
DB 62906 CTCTTAACCTTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62965
QY 1450 CAAGCTCTTGAGAGTTTAAAGCTTGATCTTATAGACCCGAAACCGTACCACTCC 1509
DB 62966 CAAGCTCTTGAGAGTTTAAAGCAATGATCTTACACACCAAAAGTTCAATCCCTC 63025
QY 1510 GATCTCTGCTGTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1569
DB 63026 AATCCACCATGTGGGCTGATCTCCATGAGGCTTGTCTTACACCTCCATGATAT 63085
QY 1570 GATTTAAAGGAAAGCGGATGATGACAGGGAACACTAGTCCCATGATGAGATGCT 1629
DB 63086 GATTTGAAAGCAAAAGGATTTACACTGCAAAATGCTTTGTTGTAAGACATTTGT 63145
QY 1630 GAGATATCAG- - -CTGGGGACTTAAGCTAGT 1658
DB 63146 GAGATCTTAGACACTATGAGGCTTAAGCTAGT 63177

RESULT 5
AF223643 1698 bp mRNA linear PLAN 16-MAY-2000
LOCUS AF223643
DEFINITION Pisum sativum xyloglucan fucosyltransferase mRNA, complete cds.
ACCESSION AF223643
VERSION AF223643.1 GI:7453578
KEYWORDS
ORGANISM Pisum sativum.
SOURCE Pisum sativum.
EUKARYOTA: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids I: Fabales: Fabaceae: Papilionoideae: Viciae: Pisum.
REFERENCE 1 (bases 1 to 1698)
AUTHORS Falk, A., Bar-Peled, M., DeRoche, A.E., Zeng, W., Perrin, R.M., Wilkerson, C., Raikhel, N.V. and Keegstra, K.
TITLE Biochemical characterization and molecular cloning of an alpha-1,2-fucosyltransferase that catalyzes the last step of cell wall xyloglucan biosynthesis in pea

JOURNAL J. Biol. Chem. 275 (20), 15082-15089 (2000)
MEDLINE 20270214
PUBMED 10747946
REFERENCE 2 (bases 1 to 1698)
AUTHORS Zeng, W. and Keegstra, K.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-2000) DOE-Plant Research Laboratory, Michigan State University, Room 110, Plant Biology Building, East Lansing, MI 48824, USA

FEATURES
Source Location/Qualifiers
1. 1698
/organism="Pisum sativum"
/db_xref="taxon:3888"
1. 1698
/function="involved in cell wall xyloglucan biosynthesis"
/note="alpha (1,2) fucosyltransferase"
/codon_start=1
/product="xyloglucan fucosyltransferase"
/protein_id="AAF62896.1"
/db_xref="GI:7453579"
/translation="MNILKRVIAIKNRGDDNNKLSDELFTLDCCTCPLTMVMAFVYSMELESVLPISLVLRDPPSDAISITTTTROLNGLSDDPDSVELNDKLGGLADBFDEKSLSTRQSAIFPKGSGKPSYLSLRVEARHQCGPTESYNTKVELGSGQSESDCKYVWISFGLGNRLTLVSAFLALDRVLDVDPGVMDLFCPEFPDASWFPDPPLNSHLNNFNOESNOCKILKTSITNSYSPFVYHLAHDYDHDKLFECDEOLFONVPLIMKTDNFIPLPSPFEGELNDLFEKKEVHFLGRVLLHPNTNNWGLVRYDAYLAKVDEKIGIQRVEDPDGPRFOHVDVACITLKESTLPVNRQNTNNSSGTPEKSAVLITLSSSGTPEKVRQMTMEFPETEVGIITPSHEGYQDQKQFHNOKAMAEVYLILSDVLTIVTSWSPFVAGGLGKWLKPENRTAPNPPCOGAMSEDFHAPPPYDCKAKKRGDTGALVPHVRHCEDMISGLKLVDN"

BASE COUNT 480 a 357 c 340 g 521 t
ORIGIN

Query Match 38.7% Score 642.6 DB 8 Length 1698:
Best Local Similarity 66.3% Pred. No. 3.8e-153;
Matches 924; Conservative 0; Mismatches 469; Indels 0; Gaps 0;

QY 267 TGATTAAGCTTTCGAGAGGCTACTGCTTCTGTTGATGAAGATCTTGCCTTAGTAG 326
DB 297 TGATTAAGCTTTCGAGAGGCTACTGCTTCTGTTGATGAAGATCTTGCCTTAGTAG 356
QY 327 GTACCAATCAGTTTCATTAACCGTAAACCTTCAACAGCCATCTTATCTCATCTC 386
DB 357 GTATCAGTACGACCATTTTCGCAAGAGACTATCAGAAACCTTCTTACCTTATTTCC 416
QY 387 TAAGCTTGAACATACGAAAGACTTACAGAGGATGCTCGGCTACTGATCTTACCA 446
DB 417 TAGACTGAGAAATATGAAAGCCGACACAAACAAATGTCCTTATCAGATCTTATATA 476
QY 447 GAAAGCTTAAACAACATGATCAAGACATTTGATGATGATGATGATGATGATGATGATGAT 506
DB 477 CAAGAAGGTGAAGAAACATCGGTTCTGCTGATCTTCTGATGCTCGGATGATGATGATGAT 536
QY 507 TGTGTGATTTCTTTTAAAGGCTTAGGGAACGATTTCTTCTAGCCCTCGGTTTTCT 566
DB 537 AGTATGATTTCTTTTAAAGGCTTAGGGAACGATTTCTTCTAGCCCTCGGTTTTCT 596
QY 567 TTAGCGGCTTTTAAAGGATAGAGTCTTCTTGAACGAGGAAGAAAGATGATGATGAT 626
DB 597 TTAGCGGCTTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 656
QY 627 CTTTTCGAGCGCTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 686
DB 657 TTTTGTGATGACATTTTCTGACAGCTTCTGCTGTTGTCCCTCGATTTTCTCTTAAACAG 716
QY 687 TCAGTTTGAATGATTAATCAAGAAATCATCTCGTTGTTATGATGATGATGATGATGAT 746
DB 717 TCATTTGAATTAATTTCAACAGAGATCGAATCAGTCAAGGGAATAATCAAGAAACAA 776
QY 747 GGTGATGATGATGAGGAACCTTGTCTCATCTTATCTTCACTCTTGTTCATGATGATGAT 806
DB 777 ATCAATCAACAATTTCACTGTCATCTTTGTTGTTATCTTCAATCAAGCCATGATGATGAT 836


```
mRNA
complement(join(<5584..5811,5909..6433,6640..6951,
7154..>7297))
/gene="At2g15460"
CDS
complement(join(5584..5811,5909..6433,6640..6951,
7154..7297))
/gene="At2g15460"
/codon.start=1
/product="putative polygalacturonase"
/protein_id="AA015318.1"
/db_xref="GI:20197930"
/translacion="MACISFAFKALCFLEFIIVASRSTVRPKVFNQVRHGAQPDGKT
DNANAFPSIMSRACKRISGSSKIYVPGTFYLGVEVFGCKNIEFIIDGTLAPN
PSIKODTYNFRYINLISGSGTLDOGKOSPHNDCHTNPCKPLAMTGFAPN
NSNIKDTLSNKMKGHNFEFVHHFNITGVTITAPGSDPMTDGIKMGCSNIOISDN
IGTGDICATLSTGTNLNINNVNCGPBGHGSVGSIGKSKDEKDYDLIVDVFNGTS
DGIRIKTWESSASKILVSNFVENIOMIDVGKPINIDOKCPHPCEHERGESHVQI
QNLKLNITGTSNKNKAVNLDQSKIFPCKNVELIDINIKONGVKGSGSTSVCENVDF
ARGMPEPRIV"
5859..5880
/rpt_family="(TAAA)n"
repeat_region
7691..7733
/rpt_family="AT_rich"
repeat_region
complement(8037..8133)
/rpt_family="AT_rich"
repeat_region
8188..8246
/rpt_family="AT_rich"
repeat_region
8912..8934
/rpt_family="AT_rich"
repeat_region
complement(8975..10695)
/rpt_family="AT_rich"
gene
/gene="At2g15450"
/note="F26H6.3"
complement(join(<8975..9208,9306..9830,10038..10349,
10552..>10695))
/gene="At2g15450"
CDS
complement(join(8975..9208,9306..9830,10038..10349,
10552..10695))
/gene="At2g15450"
/codon.start=1
/product="putative polygalacturonase"
/protein_id="A0D2279.1"
/db_xref="GI:4544368"
/translacion="MTCISFAFNALCFLEFIIVASRSTVRPKVFNQVRHGAQPDGKT
DNANAFPSIMSRACKRISGSSKIYVPGTFYLGVEVFGCKNIEFIIDGTLAPN
PSIKODTYNFRYINLISGSGTLDOGKOSPHNDCHTNPCKPLAMTGFAPN
NSNIKDTLSNKMKGHNFEFVHHFNITGVTITAPGSDPMTDGIKMGCSNIOISDN
IGTGDICATLSTGTNLNINNVNCGPBGHGSVGSIGKSKDEKDYDLIVDVFNGTS
DGIRIKTWESSASKILVSNFVENIOMIDVGKPINIDOKCPHPCEHERGESHVQI
QNLKLNITGTSNKNKAVNLDQSKIFPCKNVELIDINIKONGVKGSGSTSVCENVDF
ARGMPEPRIV"
9256..9277
/rpt_family="(TAAA)n"
repeat_region
complement(11426..11507)
/rpt_family="AT_rich"
repeat_region
11548..11635
/rpt_family="AT_rich"
repeat_region
complement(12239..12313)
/rpt_family="(CAAA)n"
repeat_region
complement(12395..12431)
/rpt_family="AT_rich"
repeat_region
12862..12890
/rpt_family="AT_rich"
repeat_region
12922..12964
/rpt_family="AT_rich"
repeat_region
13047..13128
/rpt_family="AT_rich"
repeat_region
13400..14618
/gene="At2g15440"
/note="F26H6.4: predicted by genefinder.supported by full
length cDNA: Ceres:32729"
mRNA
/gene="At2g15440"
13400..14618
CDS
13450..14439
```

```
/gene="At2g15440"
/codon.start=1
/product="expressed protein"
/protein_id="A0D2280.1"
/db_xref="GI:4544369"
/translacion="MKSNTNTNLLIFHHNSPLTSSPSAAAHNRFLFFLEFSPFT
LTFSEFLSSSIHSTTSSSLSSSSSLPPVYALALHYTSSSPSTSMFLEST
ISNTIHSHPACNLLIRGLHESILMSINPGRTVVDSPYSVSKFEQSNPGVEAY
EYVSTKVSQAKRLIGYKTRPECRPYONLFSCKIGINDLPFVYEIDMDYLIDG
PRGIASDSPGMRAPITFSAVIAKSKIDFTKTKTDVLVHEGRKIEKRVSEELCEEN
LIEVGDIGHVMAAAAEERSYGGFCFRNSTRLSEPTAVSVDGEKRVGDVE"
complement(13350..13679)
/rpt_family="(GCA)n"
repeat_region
complement(14991..15034)
/rpt_family="AT_rich"
repeat_region
complement(16086..16196)
/rpt_family="AT_rich"
repeat_region
16533..16571
/rpt_family="AT_rich"
repeat_region
complement(18723..18757)
/rpt_family="AT_rich"
repeat_region
complement(18897..18937)
/rpt_family="AT_rich"
repeat_region
18940..18991
/rpt_family="AT_rich"
repeat_region
19685..19758
/rpt_family="(TAAA)n"
repeat_region
/rpt_family="Rf:g111008031emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"
19685..19757
/rpt_family="Rf:g111008031emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"
20089..20135
/rpt_family="AT_rich"
repeat_region
complement(20944..20991)
/rpt_family="AT_rich"
repeat_region
complement(21056..21089)
/rpt_family="AT_rich"
repeat_region
21180..21230
/rpt_family="AT_rich"
repeat_region
complement(21284..21309)
/rpt_family="AT_rich"
repeat_region
complement(21445..21487)
repeat_region

Query Match 36.5%; Score 607.4; DB 8; Length 89301;
Best Local Similarity 65.0%; Pred. No. 3.4e-144;
Matches 969; Conservative 0; Mismatches 501; Indels 21; Gaps 4;

OY 175 TCAATGATCTTTACCAACACCACCTGATTCATTCGATTAATGATTAAGGTTGCGCAAGCT 234
Db 47372 TCCATGTTTGTGATTCATTTGTTACCTCAAGAAATTCCTTTTACCTTTTGTGTT 47431
OY 235 AGAGTTCTCGACCGCGGAGTTTCCCAAAATTCGATTAAGCTTTCGAGGCGCTACTGCT 294
Db 47432 ATATGTTGAGAGATTCGAAACACCGGGCGGATAGACTTAATGAGAGCGGCTTTAAC 47491
OY 295 TCTGCTTTGATGAAGATTTGCTTCTAGTAGTAC---AATCAGTTTCATTCCTGTA 351
Db 47492 GCAGATTTGATGAAGGTTCTGCTGATGATGATCAATAAACCTTCTGTATGCGCAAG 47551
OY 352 CCTTACCTTCAAGCCATCTTCTATCTCATCTCTAAGCTTAGAACCTAGCAAAAGCTT 411
Db 47552 CCTTACCTTCAAGCCGCTGTAATATCTTCTGCAAGCTTAGAAGCTATGAGATGCTT 47611
OY 412 CACAAGCATGTGTCGCGGTACTGATCTTACAGAAAGCTCTAAACCACTTGATCAA 471
Db 47612 CACAAGCTTGCGTCGAGGCAAAAGCTTACAGGAAGCAAGAACGATCTTGTGTCAT 47671
OY 472 GAACATATGATGCT-----GATGCTGATGCAATATATGTTGTGCTGATTTCT 519
Db 47672 GATGGAATTAATATGCAAGCAAGCAATAGATGATGATGCTGATGCTGCTGCT 47731
```

```

OY 520 TTACCGGCTTAGGAACAGATCTTCTCTAGCCCTGGTTTTCTTACGGCCTTTA 579
Db 47732 GATTACGGGGCTTGGAAACCGCATCTCTCTCTCTCTCTCTCTCTCTCTG 47791
OY 580 ACGATAGAGCTCTGCTGTGACCGGAGGAAGACATGATGATCTCTTGGCAGCG 639
Db 47792 ACTGATAGATCATCTCTGTGACCAACCGCAGATATGATCTCTTATGCGAGCA 47851
OY 640 TTTCTCGTATGCTGCTGCTACCTTTGATTTCCCTATGCTATCATGATTTGATGA 699
Db 47852 TTTCCAGTACTCATGCTGTCTCTCTGCACTTCCATGATGAATATCTGATGA 47911
OY 700 TTAATCAAGATCATCTGCTGTATGATATGATGAGATCAGTGTATGATCT 759
Db 47912 TACCACAGGAGTACTCTGCTGTGATGAGACATGTTGAAATCATTTCCATCACTG 47971
OY 760 GAGGGAATCTTGTCTCATCTTATCTCATCTGTTGATGATGATGATGATGATG 819
Db 47972 ACTTCATCTCCCGCATCTATATATGCTATACCTTCATGATTCAGGAGATGATGAG 48031
OY 820 ATGTTCTGCTGAGAGAGACCAACATTCATCGGAAAGTCCCTGGTGTATGTTAAA 879
Db 48032 ATGTTCTGCTGCAAAAGATCAAAAGTTGATGATGACAAAGTCCCTGGTGTATGATGA 48091
OY 880 ACAGACATTAATCTTGTCTCATCTCTGCTGATTAATACGGGTTGATGATGATGATG 939
Db 48092 GCCATGTTACTTGTCTCATCTTATGATGATGATGATGATGATGATGATGATGATG 940
OY 940 AAGCATTCCTCCAGAGAAAGCACTGCTTTCATCATCTAGTATGATGATGATGATGATG 999
Db 48152 AAGCATTCCTCCAGAGAAAGCACTGCTTTCATCATCTAGTATGATGATGATGATGATG 999
OY 1000 ACTAACAGATATGAGGCTTGTCTCATGATATCTAGATCATCTATGATGATGATGATG 1059
Db 48212 AAAATCAAGTTTGGATATGCTCTACTAGTATCATCTATGATGATGATGATGATGATG 1059
OY 1060 GAGAGATTTGGATCAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1119
Db 48212 GAGAGATTTGGATCAAGTATGATGATGATGATGATGATGATGATGATGATGATG 1119
OY 1120 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1179
Db 48332 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1179
OY 1180 GTGAGAGATCTGCCATGT---TAATACCCCAACAAAGCGCTGTTGCTACACT 1236
Db 48392 GAAGATCAAAAGTCAACATATGATGATGATGATGATGATGATGATGATGATGATG 1236
OY 1237 TTGAACGCGGCTTACGCGGAGAACTTAAGATGATGATGATGATGATGATGATGATG 1296
Db 48452 TTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1296
OY 1297 GAGGAATCTCTGCTGCTCATGACGAGCAAGAGTATGATGATGATGATGATGATGATG 1356
Db 48512 GAGGAATCTCTGCTGCTCATGACGAGCAAGAGTATGATGATGATGATGATGATGATG 1356
OY 1357 ATGCAATATGCAAGAGCTCTGCGGAATGATCTTTGATGATGATGATGATGATGATG 1416
Db 48572 GTTCAAGACCAAAAGCGCTCTGCTGATGATGATGATGATGATGATGATGATGATG 1416
OY 1417 ACAAGTCTGCTTACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1466
Db 48632 GCAAGCTCAAGTCAAACTTGGATGATGATGATGATGATGATGATGATGATGATG 1466
OY 1477 ATACTCTATAGACCGGAACCGTATCAACTCGATCTTCTGCTGCTGCTGCTGCTG 1536
Db 48692 TTACTTTATCTGCAAAAGTATCAAAAGCTCTGATCCGCTGCTGCTGCTGCTGCTG 1536
OY 1537 ATGAGCCTTCTTCACTGCTCTCAATTCATGATGATGATGATGATGATGATGATGATG 1593
Db 48752 ATGAGCCTTCTTCACTGCTCTCAATTCATGATGATGATGATGATGATGATGATGATG 1593
OY 1594 GACACGGGAACACATGATCTCTCATGATGATGATGATGATGATGATGATGATGATG 1644

```

```

Db 48812 GAATCGGGGAAGTGTCTCTTTGTTAGATGATGAGATATATGAGGG 48862

RESULT 7
AX412254
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 433 a 307 c 319 g 453 t
ORIGIN

Query Match
Best Local Similarity 36.5%; Score 606; DB 6; Length 1512;
Matches 967; Conservative 0; Mismatches 500; Indels 21; Gaps 4;

OY 178 ATGATCTTTTCAACCAACCCATGATCAATCGATTAATGAGGTTGCGCAACCTAGA 237
Db 1 ATGTTGTTGATTCATCTGTTAGCTCAAGAAATTCCTTTTACCTTTTGTGTATA 60
OY 238 GTTCTGACCCCGGAGTTTCCCAATCTGATATGACTTCTGAGAGGCTACTGCTCT 297
Db 61 TGTTCACACCAATCCCAACACCGGCGGAGTATGATGAGAGGCTTTTAAACGCA 120
OY 298 GATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 354
Db 121 GATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 354
OY 355 TCACCTTACAGCCATCTTCTTATCTCATCTGATGATGATGATGATGATGATGATGATG 414
Db 181 TCACCTTACAGCCATCTTCTTATCTCATCTGATGATGATGATGATGATGATGATGATG 414
OY 415 AAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 474
Db 241 AAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 474
OY 475 CATATTTGATGCT-----GATGATGATGATGATGATGATGATGATGATGATGATG 300
Db 301 GAGAAATTTATGCAAGCAAAATTCAGATGATGATGATGATGATGATGATGATGATG 360
OY 523 ACGGCTTAGGGAACAGGATCTTCTCTAGCTCGGCTTTTCTTTTACGCGCTTTTAAAG 582
Db 361 TACGGCTTGGAAACCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
OY 583 GATGAGTCTTGTCTGTTGACGAGGAGAAAGATGATGATGATGATGATGATGATGATG 642
Db 421 GATGAGTCTTGTCTGTTGACGAGGAGAAAGATGATGATGATGATGATGATGATGATG 642
OY 643 CTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Db 481 CCAAGTACTTCAAGTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 702
OY 703 AATCAAGATCATCTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 762

```


D	541	CACAAGGATACCTCGTGGTTACGGAACANGTGAAATTCATTCCATCAACTCGACT	600
Q	763	GGAACTTGTCTCATCTTTATCTTCATCTTGTTCATGATATGAGATCATGATPAAGTG	822
D	601	TCATTCGCCGCACATATATATATGATCAACCTTCATGATTCAGAGGATGATATGAATG	660
Q	823	TTCTTCTGTGAAGAGACCAACAATCATCTGCGGAAGCCCTGGTGTATGTTAAACA	882
D	661	TTCTTCTGCCAAAAGATCAAAAGTTGATTCACAAAGTCCCTGGTGTATTTAGAGCC	720
Q	883	GACAACTACTTGTGTCATCTGTGNGTTAAATACCGGTTTTGAGATGATCAACAAAG	942
D	721	AATGTTACTTGTGTTCCATCTTTATGTTTAAATCCAACTTTCCAAACGAACTAACGAAG	780
Q	943	CTATTTCCACAGAAAGCAGCTGTCTTTTCATCATCTAGTAGTATCTTTTCACCAACT	1002
D	781	CTGTTTCCGCGAAAGAAACCGTGTTCACCACTGGGTGGTATCTTTTCACCCCTAAA	840
Q	1003	AACCAAGATATGGGCTTAGTACATAGATACATACCAAGCTTACTATATGCCATGGCATAG	1062
D	841	AATCAAGTTGTGGATATCGTCACTAGTACATCCATGTACCTATATCCAAAGCAATAG	900
Q	1063	AAGTTGGGATTCAAAGTAAGTTTTCATTAAGACCCGGCTCCATTTACAGATGTGATG	1122
D	901	AGACTCGGATTCAAATTCGGGTTTTTTCGCATCAAGTGGATATACCAACAGTCTATG	960
Q	1123	GATCAGATTTTCATCTTGTACTCAAAAAGAGAACTTCTACTGAAGTAGACACTAGTG	1182
D	961	GACCAAGCATATCTCTCGCACACAAAGAGAGAACTTTGCTGAATTTGGTACACAGAA	1020
Q	1183	GAGAGATCTGCCCATGT---TAATACCCCAACACAAAGCCGTGCTTGCATCTTTG	1239
D	1021	GAATCAAAAAGTCAACATATGCAACATTTCCAAAAGTAAGATGTTCTGTCTACATCTTTG	1080
Q	1240	AACCGGGTTTACGGCGAGAACCTTAAGAGTATGTATTTGGAAATATCCGACATCACTGGA	1299
D	1081	TTCTCTGAGTCTCTAGAAACTAGAGAACATGTTTTTCAGACCGAGCAAAATATGACGGGA	1140
Q	1300	GAATTCATCGGTGTTTCATCAGCCGAGCCAGAGAGGTTATCAGCAGACCGAAAAAAGATG	1359
D	1141	GAGATTTATTAAGTATATACGCCAAGTGAAGAAAGGTATCAACAAACAGACAAGAGTT	1200
Q	1360	CATATATGGCAAGGCTTTGGGGAATGTATCTTTGATTTGACGATATATCTTGTGACA	1419
D	1201	CACGACCAAAAGGCGTTGTGAGATGTATCTTTTGACCTTAAACGATPAACATTTGTGCA	1260
Q	1420	AGTCTGTGTTACATTTGGATATGTAGCTCAAGGCTCTTGAGGTTTAAAGCCTTGATA	1479
D	1261	AGCTCAAGGTCAACATTTGGATATGTTCTTATATCTTTGAGAGATTAAAGCATGTTTA	1320
Q	1480	CTCATATAGACCCGAAAAACCGTACAACTCCGATCTTCTGTGTGGGCTATGTGATG	1539
D	1321	CTTATATGCGCAAAATGATATACAAAGCTCTGATCCGCGGTGTGTGATTCACATGCGATG	1380
Q	1540	GAGCTTGTTTTCCACTGCGCTTCCATTTCTATGATTTTAAGGCAA--AAAGGCTATTGAC	1596
D	1381	GAGCTTGTTTTCTTACTCTCCAGCATCATGATGATGTGAACCTGACGCAATGGGGAAGTAA	1440
Q	1597	ACGGGAACACTAGTATCCATGTGAGCATTTGTGAGATATACAGCTGG	1644
D	1441	TCGGGGAAGATGTTCTTTTGTATAGTATTTGTGAGATATATATGGGG	1488

RESULT 8	AX412452	LOCUS	AX412452	1512 bp	DNA	linear	PAT 14-JUN-2002
DEFINITION	Sequence 216 from Patent WO022675.	ACCESSION	AX412452				
VERSION	AX412452.1	GI:21444910					
KEYWORDS	thale cress.						
SOURCE	Arabidopsis thaliana						
ORGANISM	Eutariota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						

REFERENCE	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae, eurosoids II, Brassicales; Brassicaceae; Arabidopsis.			
AUTHORS	1			
TITLE	Glazebrook,J., Wang,X., Dangl,J.L., Eulgem,T. and Zhu,T. Plant genes, the expression of which are altered by pathogen infection			
JOURNAL	Patent: WO 0222675-A 216 21-MAR-2002; Syngenta Participations AG (CH) ; UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL. (US) ; Glazebrook, Jan (US) ; Wang, Xun (US) ; Dangl, Jeffrey L. (US) ; Eulgem, Thomas (US)			
FEATURES	Location/Qualifiers			
source	1..1512			
BASE COUNT	/organism="Arabidopsis thaliana" /db_xref="taxon:3702"			
ORIGIN	433 a 307 c 319 g 453 t			
Query Match	36.5%: Score 606: DB 6: Length 1512:			
Best Local Similarity	65.0%: Pred. No. 8.5e-144:			
Matches 967: Conservative	0: Mismatches 500: Indels 21: Gaps 4:			

	Query Match	36.5%;	Score 606;	DB 6;	Length 1512;	
	Best Local Similarity	65.0%;	Pred. No. 8,56-144;			
	Matches	967;	Conservative	0;	Mismatches 500;	Indels 21; Gaps 4;
OY	178	ATGATCTTTTCACCAACACCACCTGTGATTCCAATCGAATGGATTATGGCTTCCGCCAGACTAGA	237			
Db	1	ATGTTTGTCTCATTCATTTGTGTAGCTCAAGAANTTCCTTTTGACCTTTTTTTTTTTTATA	60			
OY	238	GTTCCGCAGCGCGGAGTTTCCCNAATTCATAGCTTTCGAGGGCTACTTCCTCT	297			
Db	61	TGTCGACGACGAATCCGAACAACCGGGGGGATACATTAATGAGGGCTTTTAACGCCGA	120			
OY	298	GGTWTTGATGAAGATTTCTTGCTTAGTAGTATGC---AATCAGTTTCATTACCGTAACCT	354			
Db	121	GATTTCGATGAAGGTCTCTGCTTGATGATAGTATCTATAAACTTCTTGATTCGCAAGCCT	180			
OY	335	TCACCTTCAAAGCCATCTTCTTATCTCATCTCTAAAGCTTAGAACTACGAAAAGCTTAC	414			
Db	181	TCACCATCAAGCCGCTCGAATATCTTCTCGAAGCTTAGAAGCTATGAGATGCTTAC	240			
OY	415	AAGCATGTGGCGCGGTAGTGAATCTTACAAAGAAAGCTTAAACCACTGATCAACGA	474			
Db	241	AAACGTTTCGGGTCCAGGGACAAAGCTTACAAAGAACACAAAGCATTTAGTATGAT	300			
OY	475	CATATTGATGCT-----GATGGTAAGCAATATGTTGTGTGATTTCTTTT	522			
Db	301	GAGATTTTAAATGCAAGCAAAATCAAGATGGTGAATCCGATACGTTGTGGCTCGCTGAT	360			
OY	523	AGCGGCTTAGGGAACAGGATACCTTCTCTAGCTCCGGTTTTTCTTAGCGGCTTTAAG	582			
Db	361	TACGGGCTTTGGAACCGCACTACTCTCTGTCTGTCTGTCTCTCTCTCTCTCTCTTGGACT	420			
OY	583	GATAGAGTCTTGCTTGTGTGACCGAGGGAAGACATGATGATCTTTTGGAGCGGTTT	642			
Db	421	GATAGAAATCATTTCTGTGTGACACCGCAAGGATATTTGGTGAATCTTTATGCAAGCATTT	480			
OY	643	CTCGGTATGCGVGGTSCATCTTTAGATTTCCCTATGACTGATCAAGTTTGGATTA	702			
Db	481	CCAGGTACTTCATGTTGCTCTCTCCATCTCCACATTTCCATTGATGAATAATATGATGATAC	540			
OY	703	AATCAAGAATCATCTCGTTTATATGATATATATGGAAGAATCAAGTATGATGATCTGAG	762			
Db	541	CACAAGGATACTCTCGTTGTTCGGAACAAAGTTGGAAATCATTCATCAACTCGACT	600			
OY	763	GGAATTTGTTCATCTTATCTTATCTTGTCTCAATGATATGAGATCATGATTAAGTG	822			
Db	601	TCATTTCCCGCACATCTATATATGATTAACCTTCACTTGATTCGAAGGATATGATTAAGTG	660			
OY	823	TTCTTCTGTGAAGAGACCAAAATTCATTCGGGAAGTCCCTGGTGAATGTTAAAAACA	882			
Db	661	TTCTTCTGCCAAAAGGATCAAAAGTTTGATGATCAAAAGTCCCTGGTGAATTTTGAAGCC	720			
OY	883	GACATTTACTTTGTCCATCTCTGTGGTTAATACGGGTTTCGATGATGAATCAACAG	942			
Db	721	AATGTTTACTTTGTTCATCTTATAGTTTATATTCACATCTTCCAAACGCAATTAACGAG	780			

FEATURES	source
LOCUS	AX412825
DEFINITION	Sequence 589 from Patent W00222675.
ACCESSION	AX412825
VERSION	AX412825.1
KEYWORDS	GI:21445283
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
AUTHORS	Glazebrook, J., Wang, X., Dangl, J. L., Eulgem, T. and Zhu, T.
TITLE	Plant genes, the expression of which are altered by pathogen infection
JOURNAL	Patent: WO 0222675-A 589 21-MAR-2002; SynGene Participations AG (CH) ; UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, (US) ; Glazebrook, Jan (US) ; Wang, Xun (US) ; Dangl, Jeffrey L. (US) ; Eulgem, Thomas (US)
FEATURES	Location/Qualifiers
source	1..1512
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"

Db 961 GACCAAGTCATATCTCGACACACAAAGAGAAACTTTTGCTGAAATGCTACACAGAA 1020
Oy 1183 GAGAGATCTGGCAGTGT---TAATACCCCAACCAAGCGCTGTGCATCTTG 1239
Db 1021 GAATCAAAAGTCACATATCGACATTCACAAAAGCTAAAGCTGTTGCTGCACTTTTG 1080
Oy 1240 AACGGGGTTACGGGAGAACTTAAAGATGTATGTATGGAATATCCGACATCAACTGGA 1299
Db 1081 TCTCTGAGTACTCAAGAACTAGAGAAACATGTTTCAGAGCGGACAAATATGACGGGA 1140
Oy 1300 GAATCATCGGTGTTCATCAGCCGACCAAGAGAGTTATCAGACAGCGGAAAAAGATG 1359
Db 1141 GAGATTATTAAGATATATACCCCAAGTCAAGTATCAACAAACAGACAGAAAGGTT 1200
Oy 1360 CATTAATGCAAGCTCTTGCCGAAATGTATCTTTGAGTTGACAGATATCTTGACA 1419
Db 1201 CACGACCAAAAGCGCTTGCTGAGATGTATCTTTGAGCTTAACCGATTAACATTTGCA 1260
Oy 1420 AGTCTGTGTACATTTGATGTATGCTCAAGCTCTTGAGGTTTAAAGCTTGATA 1479
Db 1261 AGCTCAAGTCAACATTTGATGTATGCTTATAGCTTTGAGGATTAAGCCATGTTA 1320
Oy 1480 CTCTTAACACCGAAACCGTACAACTCCGATCTTCTGTGCTGCGGCTATGCTGATG 1539
Db 1321 CTATATCGCCAAATGATTAACAAAGCTCTGATCCGCGTGTGTGATCAGCTGATG 1380
Oy 1540 GAGCTTGTTCACCTGCGCTCCATTTGATGTATGATTAAGGAA---AACGGTATTGAC 1596
Db 1361 GAGCTTGTTCCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Oy 1597 ACGGACACATAGTCTCTCATGTGAGACATTTGAGATATCAGCTGG 1644
Db 1441 TCGGGGAAGTGTCTCTTTGTTGATGTATGTGAGATATATGGGG 1488

RESULT 10
AF417474 1512 bp mRNA linear PLN 27-DEC-2001
LOCUS Arabidopsis thaliana fucosyltransferase-like protein FUT4 (FUT4)
DEFINITION mRNA, complete cds.
ACCESSION AF417474
VERSION AF417474.1 GI:17980428
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE 1 (bases 1 to 1512)
AUTHORS Sarría, R., Wagner, T.A., O'Neill, M.A., Falk, A., Wilkerson, C.G., Keegstra, K. and Raikhel, N.V.
TITLE Characterization of a family of Arabidopsis genes related to xyloglucan fucosyltransferase.
JOURNAL Plant Physiol. 127 (4), 1595-1606 (2001)
MEDLINE 11743104
PUBMED 21608393
REFERENCE 2 (bases 1 to 1512)
AUTHORS Sarría, R., Keegstra, K. and Raikhel, N.V.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2001) PRL, Michigan State University, 122 Plant Biology Building, East Lansing, MI 48824, USA
FEATURES
source
1..1512
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="2"
/map="At2g15390"
1..1512
/gene="FUT4"
1..1512
/gene="FUT4"
/note="AtFUT4"
/codon_start=1

/product="fucosyltransferase-like protein FUT4"
/protein_id="AAL50623.1"
/db_xref="GI:17980421"
/translation="MCHSLAQRIILLTFPEVYCSDBSETPGDRILGILTAIDPDE
GSLSKRYHKTFLYKRPSPKPESELYVLRSEYELHRLRCGCTAAVEEARHLSDEN
YNASKSDGECRYVWMLADYGLGNRLTLASLYELALDRILIVDNKRKIDGLDCEP
PGTSMLEPLDEPLMKYADYKRYSGCTGLEHNSINSTSEPHLVMHNLHDSRSD
KMEFCODSOLDKVPMILIRANYEYPSLMEFNPOTELTKLEPOKEVPHHNGRL
FHPKNQYMDIYCTKYRHDLHLSKADRLGILFVPRPDGQYQHVVDQYISQOREKLLP
ELATQESKVINISHPKSAVLTVLSPEPSKLENNMERSEKANNTEGILIKYOPSER
YQDTRKVVHDKALAEMLVLSLTNIVASSSTGYVAYISGGLKPMILVLPNNKAP
DPPCVRSITSMPCFLPPTPHCEPDNAGTESGKVPVPRCEDIMGLDEDEL"
BASE COUNT 433 a 307 c 319 g 453 t
ORIGIN
Query Match 36.5%; Score 606; DB 8; Length 1512;
Best Local Similarity 65.0%; Pred. No. 8.5e-144;
Matches 967; Conservative 0; Mismatches 500; Indels 21; Gaps 4;
Oy 178 ATGATCTTTCACCAACACCATCTGATTCAAATCCGATATGAGTTGCGCCGACCTGA 237
Db 1 ATGTTTGTCTATTCATTGTGTAGCTCAAGAATTCCTCTTTGACCTTTTGTATTA 60
Oy 238 GTTTCGACGCGCGAGATTTCCTCAATTCGTATAGCTTTCGAGGCTACTTCTCT 297
Db 61 TGTTCAGACGAATCCGAACACCGGGGGGATACACTAATAGAGGGCTTTAAACGGA 120
Oy 298 GGTTCGATGAAGATTCCTGCTTACATAGTAC---AATCAGTTCAATACCGTAACCT 354
Db 121 GATTCGATGAAGGTTCTGCTTGTAGTATCATTAACATCTTCTGATCGCAAGCT 180
Oy 355 TCACCTTCAAGCCATCTTCTATCTCATCTAGCTTAGCACTAGCAAAAGCTTAC 414
Db 181 TCACCTTCAAGCCGCTGATATATCTGCTGAGAGCTTAGCACTATGATGCTTAC 240
Oy 415 AAGCATGTGTCGCGGACTGATCTTACAGAAGCTCTAAACAACTGATCAAGAA 474
Db 241 AAGCTTGGGTCGAGGACAAACAGCTTACAGAAGCAACAAACATCTTGTATGAT 300
Oy 475 CATATTGATGCT-----GATGCTGATGCANATATGTTGTGATCTTCTTT 522
Db 301 GAGAAATTAATGAAGCAAGCAAAATAGATGATGATCCATGATGTTGTGCTCCCTGAT 360
Oy 523 AGCGCTTAGGGAACAGATATCTCTAGCCCTGGGTTTCTTACGGCGTTTAAAG 582
Db 361 TAGGGCTTGAACCGACTACTACTCTGCTTGTGTTCTCTACGCTCTTGAAT 420
Oy 583 GATAGATCTTCTGTTGACCGAGGAAAGACATGATCTCTTTGCGAGCGCTTT 642
Db 421 GATTAATCATCTGTTGTGACACGCAAGGATATTGGATCTCTTATGCAAGCATTT 480
Oy 643 CTCGATATGCTGCTGCTACTCTTATGATTTCCCTATGATGATCAGTACGTTGATGATTA 702
Db 481 CCAGCTACTTCAATGCTGCTCTCTGATCTTCCATGATGAATATGCTATGATGATAC 540
Oy 703 AATCAAGATATATCGTTGTATGATATATATGGAAGATACAGCTATGATGATGATG 762
Db 541 CACAAGGATCTCTGTTGTATGACGAACATGTTGGAAGAAATCAATCAATCAATCGACT 600
Oy 763 GGAATCTTGTCTATCTTATCTTATCTTGTATGATGATGATGATGATGATGATG 822
Db 601 TCATTCGCGCCACATCTATATATGATTAACCTTATGATGATGATGATGATGATG 660
Oy 823 TTTCTTCTGTGAAGACCAACATTCATCGGGAAGTCCCTGCTGATGTTGATTAACA 882
Db 661 TTTCTTCTCCCAAGATCAAGATTTGATGACAAGTCCCTGCTGATGATGATGATG 720
Oy 883 GACATTTACTTGTGATCAATCTGATGATTAACCGGGTTGATGATGATGATGATGATG 942
Db 721 AATGTTTACTTGTGATCAATCTTATGATGATTAACCACTTCAACCGAATCAAGAG 780
Oy 943 CTATTCACCAAGGACGCTCTTCTATCACTAGTATGATCTTTTTCACCAACT 1002

Db 781 CTGTTCCGCAAGAAACCGGTGTTCCACCAGTGGGTGCTATCTTTTTCACCCCTAAA 840

QY 1003 AACCAAGATATGGGGCTTACTGCTAGATATACGAGAGCTTATATGCAATGGCGATGAG 1062

Db 841 AATCAAGTTTGGATATGCTAGTACAGTACATGATCTATATCCAAAGCAGATGAG 900

QY 1063 AAGATTGGGATTCAGATGAGATTTTCGANGAAGACCAGGCTTCATTTTCAGCATGTGATG 1122

Db 901 AGACTCGGATTCAGAAATTCGGGTTTTCGGGATCAGATGGATGCTATCCAAACAGCTATG 960

QY 1123 GATGAGATTCATCTTGTACGCAAAAAGCAAACTTCTACTGATGACAGACTAGTG 1182

Db 961 GACCAAGCTATATCTTCACCAAGAGAAAGAACTTTTGCTGAATGCTACACAGAA 1020

QY 1183 GAGAGATCTCGCCATGT---TAATACCCCAACACAAAGCCGCTTGTGCATCTTTG 1239

Db 1021 GAATCAAAAGTCAACATATGGAACATTCCAAAAAGTAAAGCTGTTCTTGTGACATCTTTG 1080

QY 1240 AAGCGGGTTACGGGGAACCTTAAGAGTATGTTGGAAATATCCGACATCACTGA 1299

Db 1081 TCTCTAGTACTCTAAGAAAGTAAAGACATGTTTTCAGAGCAGCAAAATATGACGGGA 1140

QY 1300 GAATATCGGTCTTCTATCGACGCAAGAGGTTATGAGAGACCCGAAAGAAAGATG 1359

Db 1141 GAGATTATTAAGTATATCGCCAAAGTGGAGAAAGTATCAACAAACAGCAAGAGGTT 1200

QY 1360 CATATGCAAGCTCTTGGGAAATGATCTTTTGATTTGACAGATATCTTGTGACA 1419

Db 1201 CACGACCAAAAGCGCTTGTGAGATGTATCTTTTGAGCTTAACGATATCATTTGTGCA 1260

QY 1420 AGGCTTGTCTCATTTGATATGTAGTCAAGGCTTTGGAGGTTTAAAGCCTTGATA 1479

Db 1261 AGCTCAAGGCTCAATTTGGATATGTTGCTTATAGCTTTGAGAGATTAAAGCCATGTTA 1320

QY 1480 CTCTATAGACCAAGAAACCTTACACTCCGATCTTCGGTGGGGGTATGCTGATG 1539

Db 1321 CTTTATCTGCAAAATGATTAACAAAGCTCCGATCCGCTGTGATGATCCAGCTGATG 1380

QY 1540 GAGCTTGTTCACATGCGCTCCATCTTATGATTTGTAAGCGAA---AAGCGGATTTGAC 1596

Db 1381 GAGCTTGTTCCTTACTCTCCGACTCATGATGTAAGCTCAGCCTAGCGGAACTGAA 1440

QY 1597 AGCGGAACCTAGTCTCATGAGACATTTGAGAGATATCACTGCG 1644

Db 1441 TCGGGCAAGTGTCTTTGTTAGTATGAGAGATATATGGGG 1488

RESULT 11

AF417475 1602 bp mRNA linear PLN 27-DEC-2001

LOCUS AF417475

DEFINITION Arabidopsis thaliana fucosyltransferase-1-like protein FUT5 (FUT5)

ACCESSION AF417475

VERSION AF417475.1 GI:17980422

KEYWORDS

SOURCE

ORGANISM Arabidopsis thaliana.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1602)

Sartia, R., Wagner, T. A., O'Neill, M. A., Paik, A., Wilkerson, C. G., Keegstra, K. and Raikhel, N. V.

Characterization of a family of Arabidopsis genes related to xyloglucan fucosyltransferase1

Plant Physiol. 127 (4), 1595-1606 (2001)

JOURNAL MEDLINE 21608393

PUBMED 11743104

REFERENCE 2 (bases 1 to 1602)

AUTHORS Sartia, R., Keegstra, K. and Raikhel, N. V.

TITLE Direct Submission

Journal Submitted (07-SEP-2001) PRL, Michigan State University, 122 plant Biology Building, East Lansing, MI 48824, USA

FEATURES

source location/Qualifiers

1..1602

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="2"

/map="AL2915370"

1..1602

/gene="FUT5"

55..1602

/note="FUT5"

/note="AtFUT5"

/codon_start=1

/product="fucosyltransferase-1-like protein FUT5"

/protein_id="AA150624.1"

/db_xref="GI:17980423"

/translation="MKVLIASFGLFSLTSLNPNFKLLDATTKYDIKEKPVDK LIGGLTADPDEGSLSRHYKFLYRSPKPSSEYLVSKSEYEMLRKCGPTEY KEALIKLSRDASESNGECRIYVAVAGGNGRLTLASVFLATLERTILINDNRD VSDLCFEFPQTSMLPLDPEMLNVTYAMGKNKRPYPCYMSKSHINSTIPHLI MHNLDSDSKLFCOKDOSLDIKVEMLVQANVTVPSLMEPRTFOTELVKLPDK ETVFHHLRILFHPFNEVMDVNDYTHAHLKADERLGIQIRFGKPDGRFKAVTDV ISGTORERLLEPFAPEESKYNISRTKLSVLVASLYPERGNTLMPFSKRSSTGE IVEVYQPSGERVQDTRKSHDOKALAEYLLISLTDNIVTSARSFTGVYSLSGLKPM LLQPTNFTTTPNPCCVRSKSMPECYLTPPSHCEADWGTNSKILPFVRCEDLIYG LKIVDEF"

BASE COUNT 481 a 331 c 321 g 469 t

ORIGIN

Query Match 36.3%; Score 604; DB 8; Length 1602;

Best Local Similarity 66.6%; Pred. No. 2,8e+13;

Matches 929; Conservative 0; Mismatches 450; Indels 15; Gaps 4;

QY 268 GATTAAGCTTTCGGAGGGCTACTCTCTGTTGATGAAGATTCCTGCTTAGTAG 327

Db 181 GATTAAGCTTATAGAGGGCTTTTAACTCGGATTTGATGAAGGTTCTTCTTAGTAG 240

QY 328 TACC---AATGATTCATTACCGTAAACCTTACACCTTACAGCCATCTTATCTCATC 384

Db 241 TATCATTAATATTTCTTGTATACCGCAGCATCCCGTAAAGCCTTCTGATATCTAGTC 300

QY 385 TCTAAGCTTAAAGCTACGAAGCTTCACAGAGATGTGCGGGTACTGAATCTTAC 444

Db 301 TCTAAGCTCAGAGCTATGAGATGCTTCACAAAGCTGTGTGCTCCAGATGAGATATTAC 360

QY 445 AAGAAGCTTAAAGAACTGATCAGAAACAT---ATTGATGTGATGCGGAATGCAAA 501

Db 361 AAGAAGCAATAGAGAACTTATGTCGTGATGATCAAGCAATCAAAATGCGAATGCGA 420

QY 502 TATGTGTGTGATTTCTTTTACCGGCTTAGAGAGATGATCTTCTAGAGCTCGGT 561

Db 421 TACATTTGATGGTGGCGAGGTTACGGGCTTGGAACAGATTAATTACTTGTCTTGT 480

QY 562 TTTCTTACGGGCTTTTACAGGATAGATCTTGTGACCGAGGGAAGACATGAT 621

Db 481 TTCCCTAGAGCTCTTGTGACGAGAGATCATTTCTTCCAAACCCAGAGATGTTAGT 540

QY 622 GATCTTTTTCGAGCCGTTCTCGGTATGCGGTGTTGCTACCTTAAATTCCTGAT 681

Db 541 GATCTTTTATTCGAGCCATTTCCAGGATCAATGATGTTGCTTCGCTTGACTTCCATG 600

QY 682 ACTGATCATGTTGAT-----GGATTAATCAAGATCATCTGCTGTGATGATATAT 735

Db 601 CTGAATATTAATCTTATGCTTGGGCTTACAAATAGGAATATCTGCTGTTTACGAAACAT 660

QY 736 GTGAAGATACAGGATGATGATGAGGGAACCTTGTCTATCTTATCTTATCTTGT 795

Db 661 TCTGAAGAAATTCATCAACACTGCATTCACCCGCGCATCTATACATGATTAACCTT 720

QY 796 CATGATATGAGATCATGATAGATGTTCTCGTGAAGAGAACCAACATCATCGGG 855

Db 721 CATATTCACAGGATGATGATAGCTGTTTGTATGCCAAAGAGATCAAGGTTATGAC 780

QY 856 AAGTCCCTTGTGATTTTAAAGCAGCAATTAATGTTCTTCCATCTCTGTGTTAATA 915

Db	141937	---	ACTATGGGAATTTTACTGAGTACGCCAGAGAGCGATGTGGAAACATCGTCGAGAAATA	141993
OY	746	AGGTGATGATACCTAGGGAACCTTGTCTC	-----ATCTTATTC	784
Db	141994	AGGTGTCGGGACAAATACAGATCAGTCTCGGCTGTCTAGGCTCCCTCCAGTGTCC		142053
OY	785	TTTCATCTGTTTCATGATATGATGAGATCAATGATAAATGTTCTTGTGTAAGGAGACAA		844
Db	142054	TCCTACCTGATGATGATACATGATTAATTCATGACACACTTTTCTTGTGCGAAGACGACCAAC		142113
OY	845	CATTCATCGGGAAATGCCCTTGGTTGATTTGTTTAAACAGACAACTTCTGTTCCATCTC		904
Db	142114	AGTTCCTCAGAAATGTGCCGTTGATTCATGAGAACCGACATGTACTTATACCATC		142173
OY	905	TGTGTTAATACCGGCTTTCGATGATGAACCTAATACAGACTATTCACAGAAAGCGACTG		964
Db	142174	TATTTCTTATTCAGTTACATTCAGAAATGAGCTCAGAGGCTATTTCTGTGAAAGATGCTG		142233
OY	965	TCCTTCATCAGTTAGATGATGATCTTTTTCACACCAATACCAATATGAGGAGCTTAGTCA		1024
Db	142234	TTTTTTCATCAGCTGGCAGGCTATCTTTTCATCGCAGCAATAGCATATGCTATTCAGTTA		142293
OY	1025	CTAGTACTACGAGACTTACTTATTCGATCGCGGATGAGAGATTTGGATTCAGTAAAG		1084
Db	142294	AAGGTTACTAAGTCTCTACCTTGGCAAAAGCTAAATAAACGTGGGAATCTGATCAGACA		142353
OY	1085	TTTTTCGATGAAG-----ACCGGGTCCATTTTACAGATGATGATGATCACA		1129
Db	142354	TATTTGAAAGAGAGGCGCATCTTGCAAAAAAGGCGCATTTTCCATATGTCTTAAACACA		142413
OY	1130	TTTTCATCTTGTACTCAAAAAGAGAAACTTCTACCTGAA-----GTAGACACACTAGTGG		1183
Db	142414	TCTCTTCATGCGCTCAGATGATAAAAGCTGCTGCCAGAAATCAGTATGAAAGATGAAGCAG		142473
OY	1184	AGAGATCTCGGCATGTTTATACCOCACAAACAAAGCCGTGCTGTGCATCTTTGAACG		1243
Db	142474	AAGCAGCAGCCGCGCCCAAGATTAACAGACATATGCTGTCTTTCACAACTCTCTGAGGT		142533
OY	1244	CGGGTTACGCGGAGAACTTAAGAAGTATGATTTTGGAAATATCCGACATCACTGAGANA		1303
Db	142534	CTGTGATCAGATGATCAGATTCAGAAAGAGTACAGTAGAGCACCCACAGTCATGGAACCA		142593
OY	1304	TCATCGTGTTCATCAGCCGAGCCAGAAAGTTATCAGCAGACCGAAAAAAGATGCATA		1363
Db	142594	GGGTAGAGATATACACGCGGAGCCATGTAGAGTACACAGATCAAGAAACAAGAAAGCACA		142653
OY	1364	ATGCGAAAGCTCTCGCGGAATGATCTTTTGAATGTGACATATATCTGTGACAAGTG		1423
Db	142654	ACATGAAAGCAGCTGCGCGAGATCTTACTACTCAGCATGACGAGCAGCTGCTCATCCACAGTG		142713
OY	1424	CTTGCTTACTATTGGATATGTAGCTCAAGCTCTTGAAGTTTAAAGCCCTTGGATACTCT		1483
Db	142714	GCTTCTCCACCTTGGGCTACACCGCGGAGGCGCTCTCGGCTCGACCGCTGGATCATGT		142773
OY	1484	ATAGACCCGAAACCGTACAACTCCCGAATCTCGTGTGATGCGGCTATGTCGATGAGAC		1543
Db	142774	TCAGTGGGAAACACACCCATGCCGAGCCGCGCTGCGCGCCCAATGTCGATCGAC		142833
OY	1544	CTTGTTTCAGCTCGCTCATTTATGATTTGAAAGCAGAAACGGGTATTGACAGCGGA		1603
Db	142834	CGTGTCTTCATCAGGCTCCCTTCTACGACTCGAAGCGCAAGGAAGCAAGCTGACGTGGGA		142893
OY	1604	CACATGTTCCCTCATGTGAGACATGTGAGGATATCAGCTGGGAGCTTAAAGCTAGTA		1659
Db	142894	AGATGTGCTCCCTGCTGCGAGCAGCTGCGAGGACGTGATGTTTGGGCGCTGGAAGTTGTA		142949

RESULT 14	AC011765/c	LOCUS	DEFINITION
AC011765	134402 bp	DNA	linear
Arabidopsis thaliana chromosome 1 BAC F1M20 genomic sequence, complete sequence.			

ACCESSION	AC011765
VERSION	AC011765.6
KEYWORDS	GI:12324786
SOURCE	HTE.
ORGANISM	Arabidopsis thaliana. Arabidopsis thaliana.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopses.
AUTHORS	Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Cressy,T.H., Haas,B.J., Wu,D., Maitl,R., Rensing,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
TITLE	Arabidopsis thaliana chromosome 1 BAC F1M20 genomic sequence
JOURNL	Unpublished
REFERENCE	2 (bases 1 to 134402)
AUTHORS	Lin,X. and Kaul,S.
TITLE	Direct Submission
JOURNL	Submitted (14-OCT-1999) The Institute for Genomic Research, 9712
REFERENCE	Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
AUTHORS	3 (bases 1 to 134402)
TITLE	Town,C.D. and Kaul,S.
JOURNL	Direct Submission
COMMENT	Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org On Jan 19, 2001 this sequence version replaced gi:12280786. Address all correspondence to:ateltgr.org

BAC clone FLM20 is from *Arabidopsis thaliana* chromosome 1. The orientation of the sequence is from sp6 to 77 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHMM (Mark Borodovsky, http://www.tigr.org/softab/glimmer_hem/glimmer.html, and GeneSplicer (Mhaaela Perlea and Steven Salzberg, <http://contact.meritlab.org>), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tql.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arjan Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES
source

```

/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="FIM20"
complement(1655..3429)
/gene="FIM20.1"
/note="similar to p58 protein kinase GB:AB59449 [Homo sapiens]: contains pfam profile: PF00069 Eukaryotic protein kinase domain"
complement(join(1655..1777,1908..2135,2258..2581,2679..2963,3052..3429))
/gene="FIM20.1"
complement(join(1655..1777,1908..2135,2258..2581,2681..2963,3052..3429))

```

```

/codon_start=1
/codon_end=1
/product="putative protein kinase: 3429-1655"
/protein_id="AAG52349.1"
/db_xref="GI:12324787"
/translation="MGCVVSSKQTFVSYPALIDHGYKRNENEGSSGRIIVEPDPPT

```


D 25427 GGGAGATCGGCTGACGCGTGAAGTTTCGGTGTGATATGTAATGTTGAATTCCAA 25368
OY 522 TAGCGGCTTAGGAACAGGATACCTTCAGCCGCTTTCAGGCTTTTAAAC 581
D 25367 TGGGATCTTGGGAATGAGATGCTGAGCTGAGCTTTCAGCTTCTTATGCTCTTAAAC 25308
OY 582 GGATAGAGCTTCTGCTGCTTGTGACCGAGGAAAGACATGATGATCTCTTTGCGAGCCGT 641
D 25307 AAATAGGTTTACTTGTGCAACAGAGTGTGACATGGCTGATCTTTCTGCGAGCCATT 25248
OY 642 TCTGGATATGCTGCTGCTTGTGACCTTATGATTTCCCTATGCTGATCATGTTGATGATT 701
D 25247 TCCAAACACTGCTGCTTCTTCCCGAGAGTTTCCGCTCAACGCCACTTCAACG---- 25192
OY 702 AAATCAAGATCATCTGCTGTTTATGATATATGTTGAAGATCAGGATGATGATCTGA 761
D 25191 -----AGCAGTCTCTTACGCCAATTC 25170
OY 762 GGGAACTTGTCTCATCTTATCTTCACTTGTTCATGATGATGATGATGATGATGAT 821
D 25169 TGGCAACCCGATGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 25110
OY 822 GTTCTTCTGGAAGAGACCAACATTCATCGGAAAGTCCCTGCTGATGTTTAAAC 881
D 25109 TTTCTTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 25050
OY 882 AGAATATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 941
D 25049 GGAATGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 24990
OY 942 GCTATTCACAGAAAG 1001
D 24989 GCTTCTCCCGAAAAAGATATGCTTTCATCTTAAAGTATGATCTTCTTCAACCCAC 24930
OY 1002 TAACCAAGATGAGGCTTACTGATGATGATGATGATGATGATGATGATGATGATGAT 1061
D 24929 TAATGTTGTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24870
OY 1062 GAAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1121
D 24869 AAGATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24810
OY 1122 GGATCAGATCTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1181
D 24809 AGATCAGATCTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 24750
OY 1182 GGA---GAGATCTGCGCATGTTAAATACCCCAACAAACCGCTGCTGATCTGATCTGAT 1238
D 24749 AAATCTCCCTCAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 24690
OY 1239 GAAGCGGGGTAGCGGAGAACTTAAGAGATGATGATGATGATGATGATGATGATGATGAT 1298
D 24689 TTTCTCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24630
OY 1299 AGAATCATCGGTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1358
D 24629 AGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24570
OY 1359 GCATATGAGCAAGCTCTGCGGAAATGATCTTTGAGTTGACAGATTAATCTTTG-- 1416
D 24569 GGAAGAGTAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24510
OY 1417 -ACAAGTCTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1475
D 24509 CACAGGTTTATGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 24450
OY 1476 GATACCTATGAGCCCGAAACCGTACAACTCCGATCTCTGCTGCTGCTGCTGCTGCTGCT 1535
D 24449 GGTGTGGAACAACCTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 24390
OY 1536 GATGAGCGCTTGTTCACATGCGCTCCATCTGATGATGATGATGATGATGATGATGATGAT 1579

D 24389 AATAGAGCCCTGTTCCCAAGCAGATGTTCCATGCTGTAAG 24346
RESULT 15
AF417473 1706 bp mRNA linear PLN 27-DEC-2001
LOCUS Arabidopsis thaliana fucosyltransferase-1-like protein FUT3 (FUT3)
DEFINITION
ACCESSION AF417473
VERSION AF417473.1 GI:17980418
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE 1 (bases 1 to 1706)
AUTHORS Saaria, R., Wagner, T.A., O'Neill, M.A., Falk, A., Wilkerson, C.G.,
Keegstra, K., and Raikhel, N.V.
TITLE Characterization of a family of Arabidopsis genes related to
xyloglucan fucosyltransferase
JOURNAL Plant Physiol. 127 (4), 1595-1606 (2001)
MEDLINE 21608393
PUBMED 11743104
REFERENCE 2 (bases 1 to 1706)
AUTHORS Saaria, R., Keegstra, K., and Raikhel, N.V.
TITLE Submitted (07-SEP-2001) PRL, Michigan State University, 122 Plant
JOURNAL Direct Submission
Biology Building, East Lansing, MI 48824, USA
FEATURES
source
1. 1706
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/map="At1g74420"
1. 1706
/gene="FUT3"
62. 1543
/gene="FUT3"
/note="AtFUT3"
/codon_start=1
/product="fucosyltransferase-like protein FUT3"
/protein_id="A150622.1"
/db_xref="GI:17980418"
/translation="MKRPSLSKMTKLLAVLVAVMSLSVLEDPDSDIETE
AASRVLOSRLHGDGUSEKRAQJLNILVPSFDESCLSRKESLVKESPFKOSYL
DYRLQRYEDLHRCGPFRTSYNLTLDLKGSDRGVSGCRVYVWLNISGDLNRL
SLASAPLYALITNRLFLVELGVADLFCFEPMTWFLPPEFLNHSFNDSILRS
GNPVAAYRVHVRDSDQKLFCEDSQVLEETFWLILKADSFLPLSFVSFQEL
OMLEPEEDTAEHFLSOYLPHPTNVVGLITRYVAYLAKDORIGIYIGSESGEOR
OHILDOIACGTGHRKLLPEVDKORNLPSOVLNKSRAVFISSSPGVFXSIDVYVE
NPYMGITISVHKRSYIDYOKTFRNMSKRMATITLSSDAILVTGLMSLVEVAH
GLGGLKRWLVLNKENGTAHECYVKANSIECSOATLPHCKD-
BASE COUNT 476 a 345 c 387 g 498 t
ORIGIN
Query Match 20.7%; Score 344.2; DB 8; Length 1706;
Best Local Similarity 56.3%; Pred. No. 66-77;
Matches 761; Conservative 0; Mismatches 533; Indels 57; Gaps 4;
OY 244 GAGCCGAGTGTTCACAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 303
D 230 GATGAGTCTTACGAG 289
OY 304 GATGAGATCTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
D 290 GATGAAGATCATGCTTACGAG 349
OY 364 AAGCATCTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 423
D 350 AAGCATCTCTCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 409
OY 424 GATCGGCTACTGATATCTTACAGAAAGCTCTAAACAACT-----TGATCAAGAA 474

D	b	410	GGACCATTCCTAGACTCCTTAATACCTTAACACTTGACAACTCAAGTCGGAGATCGCTGT	469
O	y	475	CATATTGATGCGATGCGATGCAATATGTTGTGTGGATTCTTTTACGGCTTAGGG	534
D	b	470	GACGGTGAAGTTCTTGCTTGAATATGTAATATGTTGAATTCGAATCGTATCTTTGGG	529
O	y	535	AACAGATACCTTCTTCAGCCCTCGTTTCTTACCGGCTTTTAAACGGTATAGACTTGG	594
D	b	530	AATAGATGCGTAGCTACAGCTTCAGCTTCTTCTTATGCTCTCTTAACAAATAGGTTTAA	589
O	y	555	CTTGTTCACCGAAGGAAAGACATGATGATCTCTTTTGGAGCGCTTCTCGGTATGTCG	654
D	b	550	CTTGTGCACTAGAGAGTTGATGATGGCATCTTTCTTGGCAGCCATTTTCAAAACACTACT	649
O	y	655	TGGTTGCACTTATAGATTCCCTATGACTGATCAGTATGATGATGATTAATCAAGATCA	714
D	b	650	TGGTTTCTTCCCGCAGATTTCCGCTCAACAGCAGCTTCAAGC-----	692
O	y	715	TCTCGTGTATTAGATATATGCTGAGATACAGGTATGATATCTAGGGAACCTTGTCT	774
D	b	693	-----ACGACTCTCTCTCTCAGCAATTCCTGCAACCCGAGT	727
O	y	775	CATCTTATCTCATCTTGTTCATGATTAATGAGATCATGATATAGATGTTCTTCTGTGA	834
D	b	728	GTTGCAATATGCACATATGATTTGATTCAGTGACCAACAAAGCTTTTCTTTGTGAG	787
O	y	835	GGAGACCAAACTTCATCGGGAAAGTCCCTTGGTATGTTTAAACAGACATTACTTT	894
D	b	788	GATAGTCAGTTTGTGGAGAAACCCCTGGTTGATCTTAAACCGGATAGTTTCTTT	847
O	y	895	GTTCCATCTCTGTGTTAATATCCGGGTTTCGATGATGAACATAACAGCTATTCOCAG	954
D	b	848	CTCCATCTCTCTTCTCAGTCTGCTATTCATACAGAGAACTTCAATGCTCTTCCCGAA	907
O	y	955	AAACGAGCTGCTTTCATCTACTAGTAGGATCTTTTACCCACATTCACCAAGTATGG	1014
D	b	908	AAATATCTGCTTTTCACTTTAATGACATATCTTTCACCCACTATATGTTGTCTGG	967
O	y	1015	GCGTTAGTCACTAGATATCTACGAAAGCTTACTTATGCGATCGGATGAGAAAGTTGGATT	1074
D	b	968	GGCTCTATTTACAGATCTATTAACGCAATCTTGTAAAGCTGATCAAAAGATTAAGATTC	1027
O	y	1075	CATTAAGAGTTTCGATGTAAGAACCCGGGTCATTTACATGATGATGATGATGATTTCA	1134
D	b	1028	TATATTGGGCTCTGTAGTGTGCAAAATGAACAGCTTCACAGATTTGATATCATGATTTTA	1087
O	y	1135	TCTTGTACTCAAAAAGAAACTTCTACTGTGAAGTAGACACACTAGTGA---GAGATCT	1193
D	b	1088	GCTTGTGAAGTACAGCATATAGCTGTCTCTGAAAGTTGACAAACAAAGAAATCTCCCTTCA	1147
O	y	1192	CGCCATTTTAAATACCCCAACACAAAGCCGTCCTGTGCATCTTTTGAAGCGGGTTAC	1251
D	b	1148	AGCCAAATTTTGAACCGAAATATCAAGGAGAGTTTATCTCATCTTCTTCCAGGGTAT	1207
O	y	1252	CGCGAGAACTTAAAGATATGATTGTGGAAATATCCGACATCACTGAGGAATCATCGGT	1311
D	b	1208	TTTATAGAGTATCAGGGACGATATTTGGGAAACCCACAGTATGTGGAGAGTATATACG	1267
O	y	1312	GTTTCATAGCCGACCAAGAGATATTCAGCAGACCCGAAAAAAAAGATGCATATATGGCAAA	1371
D	b	1268	GTTTCACAGCCGAGCTACAGGACTACCAAAAAACCCGAGGAACATGGAAGTAAAGAA	1327
O	y	1372	GCTCTTGGGAAATGATCTTTTGAATTTGACAGATATCTGTG---ACAAGTCCTGG	1428
D	b	1328	GCATGGCTTAGATATACCTTCTGATTTTGTGATGCGCTGTGTGTCACAGGTTTATGG	1387
O	y	1429	TCTACATTTGATATGTAGCTCAAGCTCTTGGAGGTTTAAACCTTGGATCTCATATGA	1488
D	b	1388	TCTCATCTGTGGAGGTTGTCTATGGCTTTGGAGGTTGAACCCATGGGTGTTGAACAAA	1447
O	y	1489	CCGGAACCGTACAACTCCCATCTCTGTGTGCTGGGCTATGTGATGAGACCTTGT	1548

D_b 1448 GCTAGAGATGGGACGTGCCCATGAGCCTTACTGTGTGAAGCAGATGCATTAAGAACCCTGT 1507
Q_y 1549 TTCACCTCGCCCTCCATTCTATGATTTGAAG 1579
 ||||| ||| ||| |||||
D_b 1508 TCCCAAGCAGACATTGTCCATGGCTGTAAG 1538

```
Search completed: March 5, 2003, 12:12:59
Job time : 5351 secs
```


GenCore version 5.1.4-g5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 09:59:34 ; Search time 383 Seconds
(without alignments)
9772.379 Million cell updates/sec

Title: US-10-037-311A-2
Perfect score: 1662
Sequence: 1 atggatcagatcgtacag.....gggacttaagctatgta 1662

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	604.8	36.4	1720	21 AAC50203	Arabidopsis thaliana
2	248.4	14.9	875	21 AAC34268	Arabidopsis thaliana
3	50.6	3.0	171	24 ABL6082	Corn tassels-derivate
4	43.4	2.6	7001	24 ABL33919	Human DNA for stag
5	40.6	2.4	6430	24 ABLN84122	Human chemically m
6	40	2.4	32145	22 AAK68491	Human immune/haema
7	40	2.4	32145	22 AAK68575	Human immune/haema
8	39.4	2.4	6397	24 ABL32819	Human immune syste
9	38.8	2.3	4590	22 AAK24065	Yeast AOD9604-asso

10	38.2	2.3	50000	24	ABL56202	AMEPV genome fragm	
C	11	37.8	2.3	1962	24	AAS62336	CDNA sequence #123
C	12	37.8	2.3	2839	22	AAE58335	Human GRP-binding
C	13	37.8	2.3	9516	22	AAS42103	Genomic sequence #
C	14	37.6	2.3	6392	24	ABL32685	Human immune syste
C	15	37.6	2.3	6392	24	ABL32685	Human immune syste
C	16	37	2.2	792	21	ABL34507	Human metastasis a
C	17	37	2.2	15225	19	ABLN81133	Shrimp polynucleot
C	18	37	2.2	15225	21	ABLN81133	Respiratory syncyt
C	19	37	2.2	15225	21	ABLN81133	Respiratory syncyt
C	20	37	2.2	17967	24	ABL330014	Human immune syste
C	21	36.8	2.2	15229	19	AAH182167	Human immune syste
C	22	36.8	2.2	15229	20	AAH229210	Human immune syste
C	23	36.8	2.2	15229	20	AAH352668	Human immune syste
C	24	36.6	2.2	780	21	AAC36955	Human immune syste
C	25	36.6	2.2	1022	24	ABN98444	Human immune syste
C	26	36.6	2.2	1198	22	AAK62621	Human immune syste
C	27	36.6	2.2	6277	22	AAH63222	Human immune syste
C	28	36.2	2.2	5271	21	AAH80501	Human immune syste
C	29	36.2	2.2	5355	21	AAH80505	Human immune syste
C	30	36.2	2.2	23683	24	ABL70483	Human immune syste
C	31	36.2	2.2	23683	24	ABL46232	Human immune syste
C	32	36	2.2	639	24	ABO65046	Human immune syste
C	33	36	2.2	687	21	AAC37908	Human immune syste
C	34	36	2.2	688	21	AAH40855	Human immune syste
C	35	36	2.2	6060	18	AAH47484	Human immune syste
C	36	36	2.2	7025	24	ABK40055	Human immune syste
C	37	36	2.2	7025	24	AAH63355	Human immune syste
C	38	36	2.2	9789	17	AAH1854	Human immune syste
C	39	36	2.2	10328	24	ABL338149	Human immune syste
C	40	35.8	2.2	683	22	AAH08119	Human immune syste
C	41	35.6	2.1	540	23	ABV55446	Human immune syste
C	42	35.6	2.1	6874	22	AAH54546	Human immune syste
C	43	35.6	2.1	6874	24	ABN60233	Human immune syste
C	44	35.6	2.1	6874	24	ABL70381	Human immune syste
C	45	35.6	2.1	6874	24	AAH513381	Human immune syste

ALIGNMENTS

RESULT 1	
AAC50203	
ID AAC50203 standard; DNA; 1720 BP.	
AC AAC50203:	
XX 18-OCT-2000 (first entry)	
DT 18-OCT-2000 (first entry)	
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 63957.	
XX Arabidopsis thaliana	
XX Hybridisation assay; genetic mapping; gene expression control;	
KW protein identification; signal transduction pathway;	
KM metabolic pathway; promoter; termination sequence; ss.	
OS Arabidopsis thaliana.	
PN EP1033405-A2.	
XX 06-SEP-2000.	
PD 25-FEB-2000; 2000EP-0301439.	
PF 25-FEB-1999; 9905-0121825.	
XX 05-MAR-1999; 9905-012180.	
PR 09-MAR-1999; 9905-0123548.	
PR 23-MAR-1999; 9905-0125788.	
PR 25-MAR-1999; 9905-0126264.	
PR 29-MAR-1999; 9905-0126785.	
PR 01-APR-1999; 9905-0127462.	
PR 06-APR-1999; 9905-0128234.	
PR 08-APR-1999; 9905-0128714.	
PR 16-APR-1999; 9905-0129845.	

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-APR-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134378.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147492.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158939.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR	14-OCT-1999;	9905-0159330.
PR	14-OCT-1999;	9905-0159331.
PR	14-OCT-1999;	9905-0159637.
PR	14-OCT-1999;	9905-0159638.
PR	18-OCT-1999;	9905-0159584.
PR	21-OCT-1999;	9905-0160741.
PR	21-OCT-1999;	9905-0160767.
PR	21-OCT-1999;	9905-0160766.
PR	21-OCT-1999;	9905-0160770.
PR	21-OCT-1999;	9905-0160814.
PR	21-OCT-1999;	9905-0160815.
PR	22-OCT-1999;	9905-0160980.
PR	22-OCT-1999;	9905-0160981.
PR	22-OCT-1999;	9905-0160989.
PR	25-OCT-1999;	9905-0161404.
PR	25-OCT-1999;	9905-0161405.
PR	25-OCT-1999;	9905-0161406.
PR	26-OCT-1999;	9905-0161359.
PR	26-OCT-1999;	9905-0161360.
PR	26-OCT-1999;	9905-0161361.
PR	28-OCT-1999;	9905-0161920.
PR	28-OCT-1999;	9905-0161992.
PR	28-OCT-1999;	9905-0161993.
PR	29-OCT-1999;	9905-0162142.

PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0127685.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 99US-0142803.

PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 14.9%; Score 248.4; DB 21; Length 875;
 Best Local Similarity 65.9%; Pred. No. 7.6e-63;
 Matches 398; Conservative 0; Mismatches 191; Indels 15; Gaps 2;

```

QY 242 TCAGACGCCGAGTTTCCCAATTCGTATAGCTTCGAGAGGCTACTTGGTCTGCTT 301
DB 270 TCACGACGAGATCTGAAACACCGGGGATGACTATATAGAGGCTATTAACACGAGATT 329
QY 302 TTGATGAAGATTCCTGCTTAGTAGTACG---AATCAGTTCATTACCGTAAACCTTAC 358
DB 330 TCGATGAAGGTTCTGCTAGTAGTATGATCAATAAACTTTCTTGACCGCAAGCCTTAC 389
QY 359 CTTACAGCCATCTTCTATCTCATCTCTAGCTTAGAGCTTAGAAGCTAGCAAGCTTCACAGC 418
DB 390 CATACAAACCGCTGGAATATCTTGCTCGAAGCTTAGAAGCTATGAGATGCGTCAAAAC 449
QY 419 GATGTGCTCCGGGACTGCAATCTTAAGAAAGCTTAATAACACTGATCAAGACATA 478
DB 450 GTTGGGGTCCGAGGCAAAAGCTTACAAAGAAAGCAAAACACTCTAGTCATGATGAGA 509
QY 479 TTGATGCT-----GATGSGATGCAAAATATGCTGTGGGATTTCTTTACG 526
DB 510 ATTATAAAGCAAGCAAAATCAATGATGATGATCCCGATACGTTGTGGCTTGCTGATTAACG 569
QY 527 GCTTAAGACAGATATCTTCTAGCCTCGGTTTCTTTTACGCGCTTTAAAGGATA 586
DB 570 GGCTTGGAAACGACTACTCTGCTTCTGCTTCTCTACGCTCTCTTGACTGTATA 629
QY 587 GAGTCTCTCTGTTGACCGAGGAAAGACATGATGATCTCTTTTGCAGCGCTTCTCG 646
DB 630 GAATCATCTCTGTTGACCAACCGTAAAGATATTAGATCTCTTATGACGACATTTCCAG 689
QY 647 GATGTGCTGCTGCTTACTTATGATTTCCCTATGACATGATCAGTTTATGATTAATC 706
DB 650 GTACTCATGCTGCTCTCGACTTCTGACTTTCATGATGATAAATATGCTGATGATACGACA 749

```

```

QY 707 AAGATCATCTCGTTGTTATGATATATGATGAGAGATCAGGTGATGATGAGGAA 766
DB 750 AGGATFACCTCGTTGTTACGGAACATGTTGGAATAATTCATTCACATCGACTTCAT 809
QY 767 CTTTGTCTCATCTTATATCTTCACTCTGTTTATGATGATGATGATGATGATGCT 826
DB 810 TCCCGCCCATCTATATATGATGATTAACCTTATGATTAAGGATGATGATGATGCT 869
QY 827 TCTG 830
DB 870 TCTG 873

```

RESULT 3
 ABL76082
 ID ABL76082 standard; cDNA: 171 BP.
 XX
 AC ABL76082;
 XX
 DT 14-MAY-2002 (first entry)
 XX
 DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:5456.
 XX
 KW Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPS;
 KW inheritance; characteristic; growth; development; disease resistance;
 KW environmental adaptability; quality; yield; molecular marker;
 KW multigene trait; plant breeding; corn tassel; gene; ss.
 XX
 OS Zea mays.
 XX
 UN US2001051335-A1.
 PD 13-DEC-2001.
 XX
 PF 16-APR-1999; 99US-0294093.
 XX
 PR 21-APR-1998; 98US-082567P.
 PA (LALG/) LALGUDI R V.
 PA (ITOL/) ITO L Y.
 PA (SHER/) SHERMAN B K.
 PI Laijudi RV, Ito LY, Sherman BK;
 XX
 DR WPI: 2002-163647/21.
 XX
 PT Novel purified corn tassel-derived polynucleotide useful for
 PT determining altered gene expression, to recover regulatory elements and
 PT to follow inheritance of desirable characteristics through hybrid
 PT breeding programs -
 XX
 PS Claim 1; SEQ ID 5456; 201pp; English.
 XX
 CC The present sequence describes a purified corn tassel-derived
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences
 CC encode corn tassel-derived polypeptides (CDPS). The cdps sequences (I)
 CC can be used for determining altered gene expression, to recover
 CC regulatory elements and to follow inheritance of desirable
 CC characteristics through hybrid breeding programs. (I) are also useful
 CC in the evaluation, and alteration of desired characteristics associated
 CC with growth and development, disease resistance, environmental
 CC adaptability, quality and yield, and as molecular markers for studying
 CC inheritance of multigene traits in a plant breeding program. (I) can be
 CC used to produce a tassel-specific profile of gene transcription, a
 CC transcript image, to clone regulatory elements for use in transformation
 CC vectors, to express a polypeptide, to identify, isolate or extend
 CC identical or related corn tassel nucleic acid sequences from DNA
 CC libraries, in nucleic acid hybridization or amplification technologies,
 CC as query sequences to determine homology of known sequences, as probe
 CC for use in Southern or Northern hybridization, and to identify the
 CC presence of and/or to determine the degree of similarity between two
 CC (or more) nucleic acid sequences.

XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI: 2002-130908/17.
XX
XX Novel nucleic acid useful for diagnosis and therapy of diseases
XX associated with development genes such as diabetes, comprises a
XX PT sequence of a segment of chemically pretreated DNA of genes associated
XX with development
XX
XX Claim 1; SEQ ID NO 139; 27pp; English.
XX
XX The invention relates to a nucleic acid (I) comprising a sequence at
XX least 18 bases in length of a segment of chemically pretreated DNA (II)
XX of genes associated with development selected from 87 genes listed in
XX the specification such as ACCPN, ADFN, or AFDI and comprising one of 350
XX sequences (ABN79984-ABN80333) or their complements. The invention is
XX useful for the diagnosis or therapy of diseases associated with
XX development genes, in particular disease related to homeobox containing
XX genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
XX associated with congenital heart disease, epilepsy, diseases related to
XX histone deacetylation, Currarino syndrome, diseases related with the
XX development of the brain and limb girdle muscular dystrophy and dwarfism.
XX Oligomers specific to each of the genes are useful for detecting the
XX methylation state of all CpG dinucleotides within the 350 sequences or
XX (II) and their complementary sequences, as primer oligonucleotides for
XX the amplification of the 350 sequences, (II) and/or their complements and
XX as oligomer probes for detecting the cytosine methylation state and/or
XX single nucleotide polymorphisms (SNPs).
XX Note: The sequence data for this patent did not form part of the printed
XX specification but is based on sequence information supplied to Derwent by
XX the European Patent Office.
XX
XX Sequence 6430 BP; 1334 A; 334 C; 1866 G; 2896 T; 0 other:
SQ
Query Match 2.4%; Score 40.6; DB 24; Length 6430;
Best Local Similarity 45.5%; Pred. No. 0.52;
Matches 145; Conservative 0; Mismatches 174; Indels 0; Gaps 0;
QY 561 TTTTCTTAAAGCGCTTAAAGGATAGAGTCTGCTGTGACCGAGGAAGACATGGA 620
DB 1583 TTAAGTTTATATAAATGATTCGATTCGATTTTAACTGTTTAAAGTGTAAATGCGGA 1642
QY 621 TGATCTCTTTTCGAGCCGTTCTCGGTATCGTGGTGTACCTTTGATTTCCCTAT 680
DB 1643 TTGGATTTTGGTGAATGTTTTCGGGTTTGTGTTTTTTTTTTTAAATATATAT 1702
QY 661 GACTGATCAGTTTGATGATTAATCAAGAAATCATCTGCTTATGATATATGCTGAA 740
DB 1703 AATATTTTATTTTATTTTATTTTATTTTATTTTATTTATATATATATATAT 1762
QY 741 GAATCAGGATGATGATGAGGAGACTTGTCTATCTTTATCTTCATCTGTTTCATGA 800
DB 1763 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1822
QY 801 TTATGAGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 860
DB 1823 TCGTATATATATTTTAAACGTTTGAATGTAAGAGGTTTATAGTTTTTTGGAGTTA 1882
QY 861 CCGTTGGTTGATGTTTAA 879
DB 1883 ATCGTGGTTTGTATTTAA 1901
RESULT 6
ID AAK68491/C
AAK68491 standard; DNA; 32145 BP.
XX
XX AAK68491;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23303.

XX
XX Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0228287.
XX 01-SEP-2000; 2000US-0228343.
XX 01-SEP-2000; 2000US-0228344.
XX 01-SEP-2000; 2000US-0228345.
XX 05-SEP-2000; 2000US-0228509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0232968.
XX 14-SEP-2000; 2000US-0233397.
XX 14-SEP-2000; 2000US-0233398.
XX 14-SEP-2000; 2000US-0233399.
XX 14-SEP-2000; 2000US-0233400.
XX 14-SEP-2000; 2000US-0233401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.

PR	21-SEP-2000	2000US-0234377
PR	21-SEP-2000	2000US-0234474
PR	25-SEP-2000	2000US-0234597
PR	26-SEP-2000	2000US-0234998
PR	27-SEP-2000	2000US-0235584
PR	27-SEP-2000	2000US-0235536
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236358
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236502
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0239353
PR	13-OCT-2000	2000US-0239357
PR	13-OCT-2000	2000US-0239360
PR	20-OCT-2000	2000US-0241121
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0246177
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	17-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0246920
PR	17-NOV-2000	2000US-0246921
PR	17-NOV-2000	2000US-0246922
PR	17-NOV-2000	2000US-0246923
PR	17-NOV-2000	2000US-0246924
PR	17-NOV-2000	2000US-0246925
PR	17-NOV-2000	2000US-0246926
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	01-DEC-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0251989
PR	08-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251989
PR	08-DEC-2000	2000US-0251990

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
DR WPI; 2001-483426/52.
XX
XX
PI Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PI metastasis -
XX
XX
PS Disclosure; SEQ ID NO 23303; 3071pp + Sequence Listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM2170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 other;

Query Match	2.4k:	Score 40:	DB 22:	Length 32145:
Best Local Similarity	47.9k:	Pred. NO. 1.7:		
Matches 115:	Conservative 0:	Mismatches 125:	Indels 0:	Gaps 0:
Qy 692	TTGATGGATTAATCCAGAAATCATCTCGTGTATGATATATGTGTGAAGAAATCAGTGA	751		
Db 11348	TTAAGGATTTTATTAGCCAGTTTCTATATGATGCGTATATATTTTTCATCTGTGG	11289		
Qy 752	TTGATACGAGGGAACCTTGTCTCATCTTTATCTTCATCTTGTTCATGATTAATGAGATC	811		
Db 11288	CAAAATTGACAAATACCTTTTATATACATGATGATTTGCTTTTGTGATGGAGGGGAAAC	11229		
Qy 812	ATGATTAAGATTTCTTCCTGTCGAGAGACCAAAACATCATGCGGGAAGTCCCTGGTGA	871		
Db 11228	TTGTGAGAAATACATCTGGAATAGAGATTATAGGCTCATTTGATTAATGCGCATTTATTTA	11169		
Qy 872	TTGTTAAACAGACAAATTAATCTTTGTTCATCTCTGTGGTAAACCGGGTTTTCATGATG	931		
Db 11168	TTAATTGAGGCTAAAAACCTGTAATTCCTCTTTTGTGATGAGATGAGTTCCGCTCTTG	11109		

RESULT 7

AAK68575/c

ID AAK68575 standard; DNA: 32145 BP.

XX AAK68575; XX

DT 06-NOV-2001 (first entry)

XX XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23387.

XX XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KM cytostatic; gene therapy; vaccine; metastasis; ds.

XX XX

OS Homo sapiens.

XX XX

PN WO200157182-A2.

XX XX

XX Novel recombinant entomopox virus vector useful for delivering
 PT polynucleotide encoding protein to vertebrate cell, comprises
 PT polynucleotide encoding protein operably linked with heterologous
 PT promoter sequence -
 XX Disclosure; Page 201-226; 326bp; English.
 PS
 CC The invention relates to a recombinant entomopox virus (EPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell, preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents a
 CC fragment of the genome of the genus B entomopoxvirus from *Amsacta moorei*
 CC (AmpV).
 CC
 XX
 SQ Sequence 50000 BP; 20248 A; 4709 C; 4703 G; 20340 T; 0 other;
 XX
 Query Match 2.3%; Score 38.2; DB 24; Length 50000;
 Best Local Similarity 52.1%; Pred. No. 7.1;
 Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
 QY 641 TTCTCGTAGTCGTGGTCTCTACCTTTAGATTCCCTATGACTGATTCAGTTGATGAT 700
 DB 48598 TTCTCTATTATTTTATTTATTTTATTTTATTTATTTATTTTATTTTATTTTATTTT 48657
 QY 701 TAAATCAGATCATCTCGTTGTTATGATATATGATGGAAGATCGGATTCATACG 760
 DB 48658 AACATTAATAAATTTTCCGATTAATATTTTGTAAATATTAAGATTAATCATCAAT 48717
 QY 761 AGGGAACCTTGTCTCACTTTATCTTCATCTTGTTCATGATTA 803
 DB 48718 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 48760
 RESULT 11
 AAS62336/c
 ID AAS62336 standard; cDNA; 1962 BP.
 XX
 AC AAS62336;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE cDNA sequence #123 encoding novel human secreted protein.
 XX
 KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
 KW immune deficiency disorder; blood disorder; inflammatory disorder;
 KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
 KW immunosuppressive; antineumatic; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200177291-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 29-MAR-2001; 2001WO-US10485.
 XX
 PR 06-APR-2000; 2000US-195604P.

XX
 PA (GENM) GENETICS INST INC.
 XX
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ,
 PI Gulukota K, Graham JR;
 XX
 DR WPI; 2002-010900/01.
 XX
 PT New polynucleotides encoding secreted proteins useful for treating e.g.
 PT asthma, HIV and Crohn's disease -
 XX
 PS Claim 1, Page 144; 391pp; English.
 CC The present invention relates to the isolation of novel cDNA sequences
 CC which encode human secreted proteins. The cDNA sequences have been
 CC derived from a variety of human tissues. The invention also provides
 CC a method for producing proteins from these polynucleotide sequences.
 CC The proteins are useful for identifying compounds that modulate their
 CC activity and production, and the cell is also useful for identifying
 CC compounds that modulate expression of the polynucleotide sequences
 CC encoding the secreted proteins. The sequences of the invention are
 CC useful for treating diseases such as hyperproliferative disorders
 CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
 CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
 CC sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory
 CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
 CC The polynucleotide sequences of the invention are also useful in gene
 CC therapy. AAS62214-AAS6238 represent the cDNA sequences of the
 CC invention that encode for novel human secreted proteins.
 CC
 XX
 SQ Sequence 1962 BP; 620 A; 340 C; 320 G; 682 T; 0 other;
 XX
 Query Match 2.3%; Score 37.8; DB 24; Length 1962;
 Best Local Similarity 58.4%; Pred. No. 2;
 Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 708 AGAATCATCTCGTTGTTATGATATATGATGGAAGATCGGATTCATGAGGAGAC 767
 DB 668 AGAATATCCAGGTTATATCTTTCATGATGGAACGCTTGTGACACTGAAGGAT 609
 QY 768 TTTCCTCATCTTAACTTCATCTTTCATGATTAAGATCATCATAGTAAGA 820
 DB 608 CTACTAAATCTTATGATTAAATGTTCTGTCTGTAACCTGAATAATAGA 556
 RESULT 12
 AAF58335
 ID AAF58335 standard; cDNA; 2839 BP.
 XX
 AC AAF58335;
 XX
 DT 19-APR-2001 (first entry)
 XX
 DE Human GTP-binding associated protein #35 coding sequence.
 XX
 KW Human; guanosine triphosphate binding associated protein; GTP; GAP;
 KW inflammation; AIDS; Addison's disease; anemia; arteriosclerosis; asthma;
 KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
 KW osteoporosis; psoriasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200105970-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 19-JUL-2000; 2000WO-US19698.
 XX
 PR 19-JUL-1999; 99US-0144595.
 PR 23-AUG-1999; 99US-0150460.
 PR 15-OCT-1999; 99US-0159849.
 XX
 PA (INCY-) INCYTE GENOMICS INC.

```
XX Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;
PI Reddy R, Yang J, Baughn MR, Lu DAM, Azimzal Y, Patterson C;
XX WPI: 2001-091972/10.
DR P-PSDB; AAB68535.
XX
XX New guanosine triphosphate-binding associated proteins (GBAP) and their
PT encoding nucleic acids, useful for treating and/or diagnosing diseases
PT associated with GBAP expression, such as cancer, diabetes and asthma -
XX
XX Example 5; Pages 210-211; 233pp; English.
XX
XX The present invention relates to novel human guanosine triphosphate
CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their
CC coding sequences (AAAF58301-AAF58366). The proteins and coding sequences
CC of the present invention are useful for treating a variety of disorders
CC including inflammation, AIDS, Addison's disease, anaemia,
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC psoriasis.
XX
XX Sequence 2839 BP; 894 A; 493 C; 654 G; 798 T; 0 other;
SO
Query Match 2.3%; Score 37.8; DB 22; Length 2839;
Best Local Similarity 50.8%; Pred. No. 2.4;
Matches 90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
OY 675 CCCATGACTGATCGATTGATGATTAATCAAGATCATCTGTTATGATATAT 734
DB 2512 CACAAATGATTTGGCAGTTTCTTTCATTCATTAATTAATCAATGTTTAAAGAAAA 2571
OY 735 GGTGAAGATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 794
DB 2572 GTAACCTAATTTGGGTTTTCAGGCGAGTTGATATTTGACCTTAAGCTTTT 2631
OY 795 TCATGATTATGAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 851
DB 2632 TTTTTCATTTTTCAGTTAATGCTAAGAAAAAGATTTGGGAGGTTAATTAAGATAT 2688
RESULT 13
AAS42103/C
ID AAS42103 standard; DNA; 9516 BP.
XX
XX AAS42103;
AC
XX
XX 17-DEC-2001 (first entry)
DT
XX
XX Genomic sequence #419 encoding novel human enzyme polypeptide.
DE
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155301-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01239.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184654.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
```

```
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
```

CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC	isomerase or ligases. The sequences of the invention are useful in the
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of
CC	disorders including hyperproliferative disorders (e.g. cancer),
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC	blood-related disorders (e.g. haemophilia), reproductive disorders
CC	(e.g. infertility) and infectious disorders (e.g. influenza). The
CC	polynucleotides of the invention can also be used in gene therapy.
CC	AA541685-AA542192 represent DNA sequences encoding for the novel human
CC	enzyme polypeptides of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 9516 BP; 2943 A; 1810 C; 2151 G; 2611 T; 1 other;
	Query Match 2.3%; Score 37.8; DB 22; Length 9516;
	Best Local Similarity 58.4%; Pred NO.4.2;
	Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0
Oy	708 AGAATCATCTCGTGTATGATATATGTTGAAGATCAGTGATGTACTGAGGGAC 767
Db	8392 AGAATATCCAGGTTATAACTTGATCTTTCATGAACAGCTTGTGACATGAAGAT 8333
Oy	768 TTGTGTTCATCTTATCTTCATCTGTTCATGATATGAGATCATGTATAGA 820
Db	8332 CTACTAATTCCTTTACATTTTAATTTCTTCTGTCTGAACGTGAATAATAGA 8280
XX	
RESULT 14	
ABL32685	
ID	ABL32685 standard; DNA; 6392 BP.
XX	
AC	ABL32685;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 658.
XX	
KW	Human: immune system disease; cytosine methylation; antiasthmatic;
KW	antiartherosclerotic; antiandemic; cytosolic; noctropic;
KW	antiprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	antiinflammatory; antiarthritic; antidiabetic; antipsoriatic;
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neutrophromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
XX	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI: 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation
XX	
XX	Claim 1; SEQ ID NO 658; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX
SQ Sequence 6392 BP; 1818 A; 170 C; 1388 G; 3016 T; 0 other;

Query Match 2.3%; Score 37.6; DB 24; Length 6392;
Best Local Similarity 52.6%; Pred. No. 4;

Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 669 AGATTTCCTATGACATCAGTTGATGATTAAATCAAGATCATCTCGTTGTTATGG 728

DB 3646 ATATTTTCTTATGAGATGAGAAATTTTGTGTAGTTGATTGATTAATGGTTTGTAA 3705

QY 729 ATATATGCTGAGAAATCAGTGTATGATCTGAGGACCTTGTCTCATCTTATCTTCA 788

DB 3706 TTATATAGCTGATGATTGTTTTCATTTTCGGGGCTTTTATTATTATTAGTTT 3765

QY 789 TCTTGTTCATGATTAATGAGATCATGATTAAGATGTT 824

DB 3766 TTTTATTTTAAATGAAGAAAGTAAGATTTGGTGT 3801

RESULT 15

ABL34507
ID ABL34507 standard; DNA; 6392 BP.

XX ABL34507;

XX 26-MAR-2002 (first entry)

XX Human metastasis associated gene SEQ ID NO: 60.

XX Metastasis associated gene; cytostatic; gene therapy; cancer;

XX cytosine methylation; gene; ds.

XX Homo sapiens.

XX WO200177376-A2.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03970.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI: 2002-010922/01.

XX New nucleic acid derived from chemically treated metastasis genes,
XX useful for diagnosis of cancers by analysis of cytosine methylation,
XX also for treatment -

XX Claim 1; SEQ ID NO 60; 23pp + Sequence listing; English.

XX The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention.

XX Sequence 6392 BP; 1818 A; 170 C; 1388 G; 3016 T; 0 other;

Query Match 2.3%; Score 37.6; DB 24; Length 6392;
Best Local Similarity 52.6%; Pred. No. 4;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 669 AGATTTCCTATGACATCAGTTGATGATTAAATCAAGATCATCTCGTTGTTATGG 728

DB 3646 ATATTTTCTTATGAGATGAGAAATTTTGTGTAGTTGATTGATTAATGGTTTGTAA 3705

QY 729 ATATATGCTGAGAAATCAGTGTATGATCTGAGGACCTTGTCTCATCTTATCTTCA 788

DB 3706 TTATATAGCTGATGATTGTTTTCATTTTCGGGGCTTTTATTATTATTAGTTT 3765

QY 789 TCTTGTTCATGATTAATGAGATCATGATTAAGATGTT 824

DB 3766 TTTTATTTTAAATGAAGAAAGTAAGATTTGGTGT 3801

Search completed: March 5, 2003, 10:46:31

Job time : 560 secs

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 10:35:30 ; Search time 88 Seconds
(without alignments)
5792.009 Million cell updates/sec

Title: US-10-037-311A-2

Perfect score: 1662

Sequence: 1 atggatcgaattcgtacag.....ggggacttaagctagatga 1662

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTCIS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.6	2.6	7218	1 US-08-232-463-14	Sequence 14, Appl
2	37	2.2	15225	2 US-08-892-403A-2	Sequence 2, Appl
3	35.4	2.1	2935	4 US-09-480-921B-27	Sequence 27, Appl
4	35.4	2.1	7218	1 US-08-232-463-14	Sequence 14, Appl
5	35.4	2.1	8050	4 US-09-491-362-11	Sequence 11, Appl
6	35.4	2.1	8050	4 US-09-874-562-11	Sequence 11, Appl
7	34	2.0	87350	3 US-08-781-891-79	Sequence 79, Appl
8	34	2.0	87543	4 US-09-791-211-3	Sequence 3, Appl
9	33.4	2.0	14872	1 US-08-961-527-72	Sequence 72, Appl
10	32.6	2.0	7400	1 US-07-674-852-1	Sequence 1, Appl
11	32.6	2.0	7400	4 US-08-473-185-1	Sequence 1, Appl
12	32.6	2.0	7400	4 US-09-171-387-3	Sequence 3, Appl
13	32.4	1.9	5852	4 US-09-853-768-10	Sequence 10, Appl
14	32.4	1.9	7037	4 US-09-853-768-3	Sequence 3, Appl
15	32.2	1.9	4190	2 US-08-488-706-3	Sequence 3, Appl
16	32.2	1.9	4450	3 US-08-617-860B-2	Sequence 2, Appl
17	32.2	1.9	2364	2 US-08-838-219B-5	Sequence 5, Appl
18	32.2	1.9	2364	3 US-09-233-336A-5	Sequence 5, Appl
19	32.2	1.9	2364	3 US-09-233-752A-5	Sequence 5, Appl
20	32	1.9	2364	4 US-09-402-036-5	Sequence 5, Appl
21	32	1.9	2364	4 US-09-904-226-5	Sequence 5, Appl
22	32	1.9	2612	1 US-08-471-033-31	Sequence 31, Appl
23	32	1.9	2612	2 US-08-471-044-31	Sequence 31, Appl
24	32	1.9	2612	2 US-08-463-483A-31	Sequence 31, Appl
25	32	1.9	2612	2 US-08-471-046A-31	Sequence 31, Appl
26	32	1.9	2612	2 US-08-470-566B-31	Sequence 31, Appl
27	32	1.9	2612	2 US-08-838-219B-3	Sequence 3, Appl

C 28	32	1.9	2612	2 US-08-469-334-31	Sequence 31, Appl
C 29	32	1.9	2612	3 US-09-300-529-31	Sequence 31, Appl
C 30	32	1.9	2612	3 US-09-233-336A-3	Sequence 3, Appl
C 31	32	1.9	2612	3 US-09-233-752A-3	Sequence 3, Appl
C 32	32	1.9	2612	4 US-09-402-036-3	Sequence 3, Appl
C 33	32	1.9	2612	4 US-09-904-226-3	Sequence 3, Appl
C 34	32	1.9	3472	4 US-09-678-300-14	Sequence 14, Appl
C 35	31.6	1.9	1074	4 US-09-134-001C-2054	Sequence 2054, Ap
C 36	31.6	1.9	2022	4 US-09-134-001C-1643	Sequence 1643, Ap
C 37	31.6	1.9	2746	2 US-08-576-165-3	Sequence 3, Appl
C 38	31.2	1.9	218	1 US-08-700-575-23	Sequence 23, Appl
C 39	31.2	1.9	544	4 US-09-328-111-718	Sequence 718, App
C 40	31.2	1.9	792	1 US-08-565-386-5	Sequence 5, Appl
C 41	31.2	1.9	1266	3 US-09-190-821-1	Sequence 1, Appl
C 42	31.2	1.9	1266	4 US-09-535-493-1	Sequence 1, Appl
C 43	31.2	1.9	1437	4 US-09-137-222A-4	Sequence 4, Appl
C 44	31.2	1.9	4529	1 US-08-565-386-1	Sequence 1, Appl
C 45	31.2	1.9	4739	3 US-08-685-871-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F15
; US-08-232-463-14
Query Match 2.6%; Score 43.6; DB 1; Length 7218;

CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22133-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pT99pt-F15
US-08-232-463-14

Query Match 2.1%; Score 35.4; DB 1; Length 7218;
Best Local Similarity 7.8%; Pred. No. 1.3;
Matches 33; Conservative 196; Mismatches 192; Indels 0; Gaps 0;

OY 952 CAGAAAGCGACTGCTTTCATCATCTAGTAGGATCTTTTTCACCCCACTAACCAAGTA 1011
DB 1508 CAAAAACGGCATGTAGCATCACTGTAATCTATCTATGCAAGTAGTAAAGAGATA 1449
OY 1012 TGGGGCTTAGCTAGTACATAGTACAGGCTTACTATCCGATGGGAGTGAAGATTGG 1071
DB 1448 GAAAGATTGGTACRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1389
OY 1072 ATTCAAGTAGAGTTTTCATGATGAAGACCGGCTCATTTTCAGCATGTGATGATCAT 1131
DB 1388 RRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTT 1329
OY 1132 TCATCTGTACTCAAAAAGAGAACTTCTACCTGAGTACACACTAGTGGAGAGATCT 1191
DB 1328 RRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTT 1269
OY 1192 CGCATGTATATACCCCAACACAAAGCCGCTGCTGTACATCTTTGAAACGGCGTTAC 1251
DB 1268 RRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTT 1209
OY 1252 GCGGAGAACTTAAGAGATATGGAATATCCGACATCACTGAGAAATCATCGCT 1311
DB 1208 RRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTT 1149
OY 1312 GTTCATCAGCCGAGACCAAGAGTTTATCAGACGACGAAAAAAGATGCATATGCAAA 1371
DB 1148 RRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTT 1089
OY 1372 G 1372
DB 1088 R 1088

RESULT 5
US-09-491-362-11
Sequence 11, Application US/09491362
Patent No. 6281017
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
TITLE OF INVENTION: METHODS OF USE
FILE REFERENCE: WSUR14977
CURRENT APPLICATION NUMBER: US/09/491,362
CURRENT FILING DATE: 2000-01-26
EARLIER APPLICATION NUMBER: 60/118,349
EARLIER FILING DATE: 1999-02-03
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 8050
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-491-362-11

Query Match 2.1%; Score 35.4; DB 4; Length 8050;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

OY 457 AACACCTTGATCAAGAACATATGATGATGATGATGATGATGATGATGATGATGAT 516
DB 2778 AACCAATTGAGCAAGACACATGATATCTGTGACCAACAAATATCATGTTATACGT 2837
OY 517 TCTTTAGCGGCTTAGGAGAGATGATCTTCTAGCCGCTGGTTTCTTAGCGGCTT 576
DB 2838 TATTTTAAACCGGAAATAATATATTTAGTTAGTAATGTTTCAGCAAGACCTGTGT 2897
OY 577 TTAACGATAGAGTCTTCTGTTGACCGAGGAAGAACATGATGATCTT 629
DB 2898 TTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2950

RESULT 6
US-09-874-562-11
Sequence 11, Application US/09874562
Patent No. 6420159
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
TITLE OF INVENTION: METHODS OF USE
FILE REFERENCE: WSUR17549
CURRENT APPLICATION NUMBER: US/09/874,562
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/491,362
PRIOR FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: 60/118,349
PRIOR FILING DATE: 1999-02-03
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 8050
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-874-562-11

Query Match 2.1%; Score 35.4; DB 4; Length 8050;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

OY 457 AACACCTTGATCAAGAACATATGATGATGATGATGATGATGATGATGATGATGAT 516
DB 2778 AACCAATTGAGCAAGACACATGATATCTGTGACCAACAAATATCATGTTATACGT 2837
OY 517 TCTTTAGCGGCTTAGGAGAGATGATCTTCTAGCCGCTGGTTTCTTAGCGGCTT 576
DB 2898 TATTTTAAACCGGAAATAATATATTTAGTTAGTAATGTTTCAGCAAGACCTGTGT 2897

LOCATION: 39020
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42164
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42459
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46808
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46823
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46826
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 47291
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52786
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52787
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68697
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68718
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68733
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68739
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 69785
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79134
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79198
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 86336

OTHER INFORMATION: unknown
OTHER INFORMATION:
US-09-791-211-3

Query Match 2.0%; Score 34; DB 4; Length 87543;
Best Local Similarity 53.8%; Pred. No. 13;
Matches 70; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 283 GGGCTACTGCTTGGCTTGCATGAGATCTCTGCTAGTAGGACCAATCAGTCTCAT 342
DB 5379 GTGATTTTACCTCATTTATTTGAGGATGACCTTAACATATACGTTTCTGTTCTTTTAT 5438
QY 343 TACCGTAACCTTCACTTACAGCCATCTCTTATCTCATCTTAAGCTTAGAACTAC 402
DB 5439 TAATTTAAATTTTGTCTTGGTAGAATCATCTGTCTCATTTTGGAGCAAAAGATAC 5498
QY 403 GAAAGCTTC 412
DB 5499 AAAAGCCTC 5508

RESULT 9

US-08-961-527-72
Sequence 72, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ. ID NO.: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 14872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-72

Query Match 2.0%; Score 33.4; DB 4; Length 14872;
Best Local Similarity 58.6%; Pred. No. 7.8;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1232 CATCTTTGAACGGGGTTACGGCGAGAACTTAAGAGTATGTATTTGGGAATATCCACAT 1291
DB 14327 CATCTTTGCAAGAGAGCTCGGAGATTTTAAGTAAATTTATTAAGATATGAAGAAA 14386
QY 1292 CAACGTGAGAAATCATCGGTTCATCAGCCGACCAAG 1330
111 1111111 11 1111 1 111

Db 310 TCTAAGCTTAGAAGTTACGAGAAGCTTCACAA

QY	1510	GATTCCTTCGTTGGTGGGTATGATGATGAGACCTTGTTCACGCGCTGCATTCAT	1569
Db	1450	GATCCACCGGTATTCGAATCCAGTGGATGAGACCTTGTTCACCTTCTCTCCGTCTCAT	1509
QY	1570	GATTGTAAACGAAACGGGTATTGACACGGGACACTGATCTTCATGATGAGACATTGT	1629
Db	1510	GGAATGCAACCCGAATGGGGAACCTACCGGGAGGTAGTCTTGTGTGAGGATCTGT	1569
QY	1630	GAGGAT 1635	
Db	1570	GAGGAT 1575	

RESULT 2			
US-09-938-842A-1990	/ Sequence 1990, Application US/0993842A		
Patent No. US20020160378A1	GENERAL INFORMATION:		
APPLICANT: Harper, Jeff	APPLICANT: Wang, Xun		
APPLICANT: Kreps, Joel	APPLICANT: Zhu, Tony		
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING	TITLE OF INVENTION: SAME. AND METHODS OF USE		
FILE REFERENCE: SCRIPI300-3	CURRENT APPLICATION NUMBER: US/09/938, 842A		
CURRENT FILING DATE: 2001-08-24	PRIOR APPLICATION NUMBER: US 60/227, 866		
PRIOR FILING DATE: 2000-08-24	PRIOR APPLICATION NUMBER: US 60/264, 647		
PRIOR FILING DATE: 2001-01-16	PRIOR APPLICATION NUMBER: US 60/300, 111		
PRIOR FILING DATE: 2001-06-22	NUMBER OF SEQ ID NOS: 5379		
SEQ ID NO 1990	LENGTH: 1512		
TYPE: DNA	ORGANISM: Arabidopsis thaliana		
US-09-938-842A-1990			

Query Match	36.5%, Score 606; DB 9; Length 1512;		
Best Local Similarity	65.0%; Pred. No. 2,4e-166;		
Matches 967; Conservative 0; Mismatches 500; Indels 21; Gaps 4;			

QY	178	ATGATCTTCACCAACACCCATCGATTCACAAATGGGATATAGGTTTCGCCGAAGCTAGA	237
Db	1	ATGTTTGTCTATCTATTCTTAGCTCAAGAAATCTCCTTTGACCTTTTGTGTATATA	60
QY	238	GTTCTGACGCCGAGTTTCCCAATTCGTGATAGGCTTCGAGAGGCTACTTGTCTT	297
Db	61	TGTCAGACGATCCGAAACACCGGGCGGAGTAGACATAATAGAGGGCTTTAAACGCCA	120
QY	298	GATTGTGATGAAGATCTTCGCTTAGTAGTAGAC--AATCAGTCTTACCGTAAACT	354
Db	121	GATTTCATGTAAGTCTCTGCTTAGTAGTAGATATAAACTTTCTTGTATCCAAACCT	180
QY	335	TCACCTTAAACAAGCATCTCTTATATCATCTCTAAGCTTAGAACTAGCAAGAAAGCTTCAAC	414
Db	181	TCACCATCAAGCCGCTGATATCTTGTCCGAGCTTAGAAGCTATGAGATGCTTCAAC	240
QY	415	AAGGATGTCGCGGGGTACTGATCTTACACAAAGAGCTTAAACAACCTTGATCAAGAA	474
Db	241	AAAGGTTCGCGGTCCAGGACAAAGCTTACAGGAAACAAAGCAAGCATTAATCATGAT	300
QY	475	CATATTGATGT-----GATGTGAATGCAAAATATGTGTGTGATTTCTTT	522
Db	301	GAGATTTATATGCAACCAATCAGATGCTCATTCGCCGATACGTGTGTGCTGCTGAT	360
QY	523	AGCGGCTTAAAGGAACAGATACCTTCTCTAGCCGTCGGTTTCTTTTAAAGCGGCTTTAAAG	582
Db	361	TACGGGCTTGAACCAACGACTACTCCTTGTCTTGTGTCTCTTACGCGTCTTGAT	420

```

RESULT 2
US-09-938-842A-1990
; Sequence 1990. Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT00-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1990
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1990

```



```
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20010051335A1 700356786H1
LOCATION: 65
OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-5456
```

```
Query Match
Best Local Similarity 3.0%; Score 50.6; DB 10; Length 171;
Matches 95; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
```

```
1219 GCGCTGCTTGACATCTTTGAGCGGGTTACCGGAGAGAACTAAAGATATATG 1278
1 GCGGTGCTGATCGGCTGCTTACTCGGGGACTACAGAGAGATCCCGGATGTACTAC 60
1279 GAATATCCGACATCACTGAGAAATCATCGGTTCATCAGCCGACCAAGATTAT 1338
61 GAGANCCCCACAGAGTGGGAGATCGTGGCGGTGTTCCAGCCGACCAAGAGACAG 120
1339 CAGCAGACCCGAAAAAAGATGCAATATGCGCAAGCTCTTGGCGAATGTA 1388
121 CAGCAGTACAGTCCACAGACCAACAGAGGCGCTGGCGGAGATCTA 170
```

RESULT 5

```
US-09-923-876-3543
Sequence 3543, Application US/09923876
Patent No. US20020013958A1
GENERAL INFORMATION:
APPLICANT: Laigudi, Raghunath V.
APPLICANT: Kamigaki, Laura V. (Ito)
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: FL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923,876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/298,329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO 3543
LENGTH: 228
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020013958A1 700162471H1
LOCATION: 196, 203
OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-3543
```

```
Query Match
Best Local Similarity 2.9%; Score 47.6; DB 10; Length 228;
Matches 113; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
```

```
1255 GAGAACTTAAGAGTATGTAATGGGAATATCCGACATCACTGGAGAAATCATCGGTGT 1314
5 GAGAGGCTGAGTGCAGTACTACGACCAAGGGGGAAGGTGGGAGAGGGTGGCGTG 64
1315 CAGCAGCCGAGCCAGAGATTATCAGCAGACCGGAAAAAAGTGTGATATGCGAAAGCT 1374
65 TTCACACCGTCGACAGAGAGCGCCGAGGAGACATGGCGACACAGAGAGCG 124
1375 CTTGCGGAATGATCTTTGAGTTGACAGATATCTGTGTGCAAGTCTGTTGTTACA 1434
125 CTTGGGAGATATCTACTGCTCAGCTTCTCCGAGAGACCTTCTACGTCGGGAGGTCACG 184
```

```
QY 1435 TTGGATATGATGCTCAAGGCTTGGAGGTTTAAAGCCTTGAT 1478
DB 185 TTGGGTACGTNAGCAGCNACCTCGCGGCTCCGCGCAGAT 228
```

RESULT 6

```
US-09-878-574-4288
Sequence 4288, Application US/09878574
Patent No. US2002010548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(1540)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 4288
LENGTH: 545
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(545)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: LIB3028-013-Q1-B1-C1
US-09-878-574-4288
```

```
Query Match
Best Local Similarity 2.4%; Score 39.2; DB 10; Length 545;
Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
```

```
QY 474 ACATATTGATGGATGGATGGAATGCAATATATGTTGATGATGCTTTAGCGGCTTAG 533
DB 42 ACATAGTCTCGCGGATGTCACAGNTCCGCTTGGGTTTTTTTTTTTCTTGG 101
QY 534 GAACAGATATCTTCTATGACCTCGGTTTTCTTTACGCGCTTTTACGAGATAGACTCT 593
DB 102 GTTCTCTCTCCTCTCTCTTCTCTGCTGCTGCTGCTCTCTCTCTCTCTCTCT 161
QY 594 GCTTGTGACGAGGAGAAAGATGATCTCTTTGGCAGCCGTTTCGATGTC 653
DB 162 TTGCTTCTTTTGGGGGTTTCTGTGTTTTTGTTCCTCCGGGCTTTTCTCTGCTGTG 221
QY 654 GTGTTGCTACCTTATGATTTCCCTATGACTGATCAGTTGAT 696
DB 222 GGGCTCCTTTCGTTTGTGTGCTCTGTTTTCGCTCTTTGTT 264
```

RESULT 7

```
US-09-822-830A-123/C
Sequence 123, Application US/09822830A
Patent No. US20020142952A1
GENERAL INFORMATION:
APPLICANT: Genetics Institute, Inc.
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakara
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822,830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,604
PRIOR FILING DATE: 2000-04-06
```



```
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 123
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-123
```

```
Query Match          2.3%; Score 37.8; DB 10; Length 1962;
Best Local Similarity 58.4%; Pred. No. 1.1;
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

```
QY 708 AGAATCATCTCTTATGATATATGATGGAAGATCAGTGATTAAGTGAAGAAC 767
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 668 AGAATATCCAGCTTAATATCTGATTCCTTCAATGAAGACCTTTGTGACACTGAAGAA 609
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 768 TTGTCTCATCTTATCTTCATCTTGTTCATGATTTATGAGATCATGATTAAGA 820
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 608 CTACTAATATCTTATGATTTAATGTGTCTGTGTAAGCTGAAGAAATAGA 556
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

RESULT 8

```
US-09-770-445-210/c
; Sequence 210, Application US/09770445
; Patent No. US2002023281A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 1022
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-210
```

```
Query Match          2.2%; Score 36.6; DB 10; Length 1022;
Best Local Similarity 68.0%; Pred. No. 1.8;
Matches 51; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
```

```
QY 421 TGTGCTCCGGTACTGATCTTACAGAAAGCTTTAAACCACTTATCAAGACATATT 480
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 807 TGTAGTGGAGCAATGAAATATAGAGAAAGATTAATACATCTGAGAGGAAAGATATA 748
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 481 GATGTGATGTGTA 495
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 747 GATGTGACACTGAA 733
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
RESULT 9
US-09-924-035A-534/c
; Sequence 534, Application US/09924035A
; Patent No. US20020142319A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2011US
; CURRENT APPLICATION NUMBER: US/09/924,035A
; CURRENT FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,784
; PRIOR FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 534
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-924-035A-534
```

```
Query Match          2.2%; Score 36; DB 10; Length 328;
Best Local Similarity 58.3%; Pred. No. 1.4;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 1310 GTGTTCATCAGCCGACCAAGAGTTATCAGCAGCAACCAAAAAAGATCATATGCA 1369
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 312 GATTTGTCGTCGTGAGCAACGAGTCATGATGATCTTACATCTTACGATTCCTCC 253
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1370 AAGCTCTGCGGAATGTATCTTTGAGTTTGACAGATATATCTTGGA 1417
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 252 ATGCTTAAGCGGCTGTGTCATTTGGTTTACTTATCATGTTTGA 205
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

RESULT 10

```
US-09-770-149-623/c
; Sequence 623, Application US/09770149
; Patent No. US2002005963A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 623
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)...(639)
; OTHER INFORMATION: n = A,T,C or G
US-09-970-149-623

Query Match      2.2%; Score 36; DB 10; Length 639;
Best Local Similarity 58.3%; Pred. No. 2.1;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1310 GTGTTACACACCCGACCAAGATTATTCAGCAGACCGAAAAAGATGCAATATGGCA 1369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 GAGTGTGCGGTGAGACAGAGTATGAGTCTACATCTTAAAGGATCCTGATTTCCC 241

QY 1370 AAGTCTTCGGAATATGATCTTTGAGTTGACAGATTAATCTGTGA 1417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 240 ATGCTTAAGCGCGCTGGTGCATTTGGGTTGACCTTATCATGTTTGA 193

RESULT 11
US-09-974-300-2061
; Sequence 2061, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2061
; LENGTH: 944
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2061

Query Match      2.1%; Score 35.4; DB 10; Length 944;
Best Local Similarity 53.2%; Pred. No. 3.8;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1243 GCGGGTACGCGGGAAGTAAGATGATGTATGGAAATATCCGACATCACTGAGAA 1302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GAGCTTGACCAAGAAAGTAATGATGAAACAGTTTAAAGAAAATACCGAAAACATTTGAGAAA 480

QY 1303 ATCATCGGTGTTTCATCAGCCGACCAAGATTATCAGCAGACGAAAAAAGATGATCAT 1362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 ATCTTCGCGCAAGAAAGATGTGTCAAGAAAGCCCTTGCGAAATTCATGAAAAAGTAGAG 540

QY 1363 AATGGCAAGACTTTGCGGAA 1383
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AAGCCAAACAGCTGGCGAA 561

RESULT 12
US-10-047-412A-27
; Sequence 27, Application US/10047412A
; Publication No. US20020197696A1
; GENERAL INFORMATION:
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Budziszewski, Gregory J.
; APPLICANT: Potter, Sharon L.
; APPLICANT: Wegrich, Lynette M.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-30780DIV
; CURRENT APPLICATION NUMBER: US/10/047,412A
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 27
; LENGTH: 2935
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-047-412A-27

Query Match      2.1%; Score 35.4; DB 9; Length 2935;
Best Local Similarity 50.3%; Pred. No. 7;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 457 AACCACTGATCAAGACATATTGATGGATGGATGCAATGCAATATGTTGTGGATT 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2335 AACCAATTGACAGACAGATGATCTGTGGAACAAACAAATCATGTTATACGT 2276

QY 517 TCTTTAGCGGCTTAGGACAGATACCTTCTAGCCCTGGTTTCTTACGCGCTT 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2275 TATTTTAAAAACCGGAAAAATATATATTAGTTAGTAATGTTTCAGCAGACCTGTGT 2216

QY 577 TTACGGATAGAGTCTTCTGTTGACCGAGGAAAGACATGATGATCTT 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2215 TTCAATCATGGAATGATGATATCTTGGGATGAATGACATCTCTATATCGT 2163

RESULT 14
US-09-969-373-1042/c
; Sequence 1042, Application US/09969373
; Patent No. US20020133852A1
; GENERAL INFORMATION:
; APPLICANT: Effertz, Roger J.

US-09-987-025-1/c
; Sequence 1, Application US/09987025
; Patent No. US20020108148A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert
; APPLICANT: Campos, Narcisco
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: Nucleic Acid Sequences Involved In
; FILE REFERENCE: 17142/02/US
; CURRENT APPLICATION NUMBER: US/09/987,025
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/549,787
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/129,899
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/146,461
; PRIOR FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3400
; TYPE: DNA
; ORGANISM: Arabidopsis sp
US-09-987-025-1

Query Match      2.1%; Score 35.4; DB 10; Length 3400;
Best Local Similarity 50.3%; Pred. No. 7.6;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 457 AACCACTGATCAAGACATATTGATGGATGGATGCAATGCAATATGTTGTGGATT 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2335 AACCAATTGACAGACAGATGATCTGTGGAACAAACAAATCATGTTATACGT 2276

QY 517 TCTTTAGCGGCTTAGGACAGATACCTTCTAGCCCTGGTTTCTTACGCGCTT 576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2275 TATTTTAAAAACCGGAAAAATATATATTAGTTAGTAATGTTTCAGCAGACCTGTGT 2216

QY 577 TTACGGATAGAGTCTTCTGTTGACCGAGGAAAGACATGATGATCTT 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2215 TTCAATCATGGAATGATGATATCTTGGGATGAATGACATCTCTATATCGT 2163
```

```
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1042
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-1042

Query Match
Best Local Similarity 53.2%; Score 35; DB 10; Length 229;
Matches 74; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

OY 689 AGTTGATGATTAATCAAGATCATCTGTTGTTATGATATATGAGATCAGG 748
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 AATTAAATTAATTAAGTAATAATACATTAATTAATGCTCATTAATTAATAA 121

OY 749 TGATTGACTGAGGAGACTTGTCTCATCTTATCTCATCTGTCATGATGAG 808
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 120 TTAATCATTAATTTTAAAGTTGTTTATCTCATTTGATTTGTTAATTTCTTAT 61

OY 809 ATCATGATAAGATGTTCTT 827
    | | | | | | | | | |
Db 60 ACTAATATAATATTATTATT 42

RESULT 15
US-09-733-507-3
; Sequence 3, Application US/09733507
; Patent No. US20010025379A1
; GENERAL INFORMATION:
; APPLICANT: Agriculture and Agri-Food Canada; The University of Saskatchewan
; TITLE OF INVENTION: Cyclin Dependant Kinase Inhibitors as Plant Growth
; FILE REFERENCE: 81601-3
; CURRENT APPLICATION NUMBER: US/09/733,507
; CURRENT FILING DATE: 2000-12-02
; PRIOR APPLICATION NUMBER: CA 2,256,121
; PRIOR FILING DATE: 1998-12-31
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 660
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-733-507-3

Query Match
Best Local Similarity 66.7%; Score 35; DB 10; Length 660;
Matches 50; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

OY 421 TGTGTCGCGGACTGAATCTTACAGAAAGCTCTAAACAACCTGATCAGAACATATT 480
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 TGTAGTGAAGCAATGATATAGAAGAAAGATTAATACATCTGAGAGAGAGATAAA 156

OY 481 GATGTTGATGTTGAA 495
    | | | | | | | | | |
Db 157 GATGTTGACACTGAA 171
```

Search completed: March 5, 2003, 12:37:33
Job time : 173 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:29:14 ; Search time 21 seconds
(without alignments)
2554.429 Million cell updates/sec

Title: US-10-037-311a-1

Perfect score: 3004

Sequence: 1 MDONSRRSSPIRTTGGSGS.....GTLVPHVHCEDISWGLKLV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3004	100.0	558	2 T02704	hypothetical prote
2	1706.5	56.8	526	2 A86274	F7A19.15 protein -
3	1679	55.9	539	2 T02705	hypothetical prote
4	1613.5	53.7	500	2 D86274	hypothetical prote
5	1583	52.7	537	2 B86274	F7A19.16 protein -
6	1528	50.9	503	2 D84528	hypothetical prote
7	1492	49.7	533	2 B84528	hypothetical prote
8	1411.5	47.0	440	2 H84527	hypothetical prote
9	1302	43.3	435	2 B86274	hypothetical prote
10	1124.5	37.4	562	2 A86773	hypothetical prote
11	112.5	3.7	2149	2 C96695	ribulose biphosph
12	110	3.7	529	2 T23431	hypothetical prote
13	110	3.7	658	2 A64584	hypothetical prote
14	110	3.7	1102	2 S35617	hypothetical prote
15	107	3.6	567	2 C82183	probable oridin-bi
16	106.5	3.5	80	2 A84528	catalase VC1585 (1
17	106	3.5	3844	2 A84528	hypothetical prote
18	105	3.5	381	2 T15140	asparagine/asparta
19	104.5	3.5	1583	2 T14176	hypothetical prote
20	103.5	3.4	1209	2 T16663	probable phosphati
21	102	3.4	832	2 H72278	hypothetical prote
22	100.5	3.3	409	2 E86017	alpha-mannosidase-
23	100.5	3.3	494	2 E91171	probable beta-kefo
24	97	3.2	594	2 B71893	excinnuclease ABC c
25	97	3.2	678	2 C83878	methy-accepting c
26	97	3.2	2044	2 A81180	probable peptidogl
27	97	3.2	4717	2 T41581	hypothetical colle
28	96	3.2	594	1 E64622	excinnuclease ABC c
29	96	3.2	811	2 A80680	probable dimethyl

30	95.5	3.2	835	2 T05259	probable disease r
31	95	3.2	326	4 S61652	hypothetical prote
32	95	3.2	1802	2 G71616	hypothetical prote
33	94	3.1	717	2 T27066	hypothetical prote
34	94	3.1	1333	2 S30356	hypothetical prote
35	94	3.1	1376	1 J01534	CDC25 protein homo
36	93.5	3.1	808	2 F90915	E2 glycoprotein pr
37	93.5	3.1	808	2 C85764	probable oxidoredu
38	93.5	3.1	808	2 F64914	probable oxidoredu
39	93.5	3.1	1145	2 T33606	dimethylsulfoxide
40	93	3.1	455	2 S61159	hypothetical prote
41	93	3.1	673	1 VCPV85	CDC40 protein - ye
42	93	3.1	837	1 A56681	coat protein VPI -
43	93	3.1	1338	2 T18416	penicillin amidase
44	93	3.1	1418	2 S64918	hypothetical prote
45	93	3.1	1817	2 D71606	hypothetical prote

ALIGNMENTS

RESULT 1
T02704
hypothetical protein At2g03220 (imported) - Arabidopsis thaliana
N:Alternate names: hypothetical protein T18E12.11
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02704; G84445
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.
A:Reference number: 214702
A:Accession: T02704
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-558 <ROU>
A:Cross-references: EMBL:AC005313; NID:g3548797; PID:g3548808
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84445
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <STO>
A:Cross-references: GB:AE002093; NID:g3548808; PID:AC34480.1; GSPDB:GN00139
C:Genetics:
A:Gene: T18E12.11; At2g03220
A:Map position: 2
A:Introns: 93/1
Query Match
Best Local Similarity 100.0%; Score 3004; DB 2; Length 558;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDONSRRSSPIRTTGGSGSVNFSLLQMKYLLSGTWKLTFTTCLIVSVLAFSM 60
DB 1 MDONSRRSSPIRTTGGSGSVNFSLLQMKYLLSGTWKLTFTTCLIVSVLAFSM 60
QY 1FH0HPSDSNRIMGFPAEAVLDAGVPPNVTNINSOKLLGGLASGPDSCSRRQSVHY 120
DB 1FH0HPSDSNRIMGFPAEAVLDAGVPPNVTNINSOKLLGGLASGPDSCSRRQSVHY 120
QY 121 RRPSPYKPSYLLSKLRNKKRGGPGTESYKALKOLDOEHIDGDECCYVWVWISFS 180
DB 121 RRPSPYKPSYLLSKLRNKKRGGPGTESYKALKOLDOEHIDGDECCYVWVWISFS 180
QY 181 GIGNRILSLASVFLYALLTDRLVLDVRGKMDLFCPEFLGSWLLPLDFPMTDQFDGIN 240
DB 181 GIGNRILSLASVFLYALLTDRLVLDVRGKMDLFCPEFLGSWLLPLDFPMTDQFDGIN 240

RESULT 4
D86274
hypothetical protein F7A19.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: D86274
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso
Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Hutzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-500 <STO>
A:Cross-references: GB:AE005172; NID:G5080784; PIDN:AMD39294.1; GSPDB:GN00141
A:Genetic:
A:Map position: 1

RESULT 5
B86274
F7A19.16 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: B86274
R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; White, O.; Alon, U.; Chiu, C.W.; Hughes, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,ansen, N.F.; Chung, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matti, R.; Marziano, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallidker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the Plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <STO>
A:Cross-references: GB:AE005172; NID:95080783; PIDN:AAD3293.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1

Query Match	52.7%	Score 1583;	DB 2;	Length 537;
Best Local Similarity	59.1%	Pred. No. 3.7e-120;		
Matches 317;	Conservative 71;	Mismatches 122;	Indels 26;	Gaps 8;
OY	10MKYLSGTMKLTFRFTCTCLVFSVLVAFSWIFHQHPDSNRIMGPFARVADAGVFN 88			
Db	17 LKMKLL-----LTIVFSGLIWSVYLVFS-----NDFFNOLLVATSNVSRSEFPR 63			
OY	89 VTNINSDKLLGLSAGDEDSCLSRV-QSVHYRRKSPKPPSYLISKLRNEKLHKRCG 147			
Db	64 -----DRLIGLLIADPDEGSCLSRYQOSSLYKRASPKREYLVSKRSYEKLHKRCG 117			
OY	148 PETESTYKKAALKO--DQEHIDGD--GECKYVYVWISFGLNRLISLASVETVALLTDRLV 203			
Db	118 PETDYAKKATELLIGHDDEMYASKSVGECRYIYVWVAVYGGNRLITLASVETVALLTERV 177			
OY	204 LVDRCKDDMDLCEPFGLGSMWLLPLDPFMTDFGLNDESSRCVGYMYKNQYIDPDEGLS 263			
Db	178 LVYQSKDLSIDLECEFPFGSTWMLPLFEPFLMKQIDQYNGKYSKCYGTMNLQAINSTLPR 237			
OY	264 HLYLHLVMDYGDHDMFCEGSDQTFIGKVPMLIVTDVYFVPSLWLIPGFDDELKLP 323			
Db	238 HLYLHLIHDRNDKMFCEQCKQKDSLVDKVPMLIVANVYFVPSLWLNLPFQELMKLP 297			
OY	324 KATVTHHLGRLYFHPRTNOVWGLVTRYEAYLSHADEKIGIQVRFDEDPGRFOHWMDIS 383			
Db	298 KEAVVHNHLARYLFIHPRTNOVWGLITRSYNAYSRADETLGIIQIRFSDRAGYFOHWMDV 357			
OY	364 SOTQKELLPEVDTLVSSRHVN-TPKHKAVLVYSLMNGYAEMLKSMWREYTSIGETIG 442			
Db	368 ACTRENENLPEPAAOEPRVNISSOKLKAVALYVTSLYPEYSETLKNMWERSSTGEIIE 417			
OY	443 VHPQDEGVQOQTEKKMHNGKALAEYLLSLVDNLVTSAMSTGYAAGGLGGLKPIYLPR 502			
Db	418 VYQPSGGEVQOQDKLHNOKALAEYLLSLVDNLVTSAMSTGYAHSGLGGLKPIYLLQRP 477			
OY	503 ENRTTPDPSCGRAMSEPCFHSPPYDCKAKTGIDGTGLVLPVHRHCEDI-SWGLKL 557			
Db	478 TGRTPADPDCIOGSTMDPCNLLPRSHSGCEPEMGSTGSMVPRVPRCEBRGNDGLTL 533			

[illegible][illegible]

Db 406 --QMSDVLNKKKAAQOFVKS 425

RESULT 13

hypothetical protein HP0513 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: A64584

R:Tomb, J.F., White, O., Kellavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalek, H.G., Glodek, A., McKenney, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Wathey, L., Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: A64584

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-658 <TOM>

A:Cross-references: GB:AE000565; GB:AE000511; NID:92313616; PIDN:AA007579.1; PID:9231362

Query Match 3.7%; Score 110; DB 2; Length 658;

Best Local Similarity 17.8%; Pred. No. 1.2;

Matches 102; Conservative 79; Mismatches 191; Indels 202; Gaps 23;

37 GMMKLTFFETTCCLYFSVLVAFSMFIHQPSDSNRIMGFAERVLDAAGVFPVNTINSDK 96

Db 28 GQORIT-TITLLLA-----LRNHLSEVEILKFSRKESLYLIN-SKMDK 74

Qy 97 LIGGLASGFEDSCLSRYSQSVHYKRPSPYSSYLISKLRYEKLHRCGPGTESYKA 156

Db 75 KF-RLISESKDITLSIDK-NKRKPS-EPSYKIVENFELFEKWIENDKLETITKG 130

Qy 157 LKQDQEHIDGDECKYVWIS-----FSGICNRLISLASVFLYALLDRVLLV 205

Db 131 LKLT-----MTWISLIDKGDPPQILFESMNSKIDELTQT--DLIRNYIME 175

Qy 206 DRGKMDLFCERFLGMSMLPLDEPMTDOF----- 236

Db 176 TEVEGEQEFYQYKRAMEREONETLNRFRVRYHLYTKIGIPNEKRYENAFKDYROK 235

Qy 237 ---DGLNQSSRCYGVW---KNQYIDTEGTLN-----LYLHVHYGDH 276

Db 236 GIEIDILKDLQKCYGCOIAFKKEDDKDLNKALSLVLEMDIYLLLEISDYND- 294

Qy 277 DKMFCEBDQFFIGKVPMLYKTDNYFVPSLMLIPG-----DDELNKLFP--- 322

Db 295 -----GVLKSD--FPIPIYLIIESYICRAVGLGYNISNKVFPSTK 335

Qy 323 --OKATVFHNLGRYLF-----HPTNQVMGLV-----TRYEAVLSHADKIG 362

Db 336 HIQKDEYKSLKANHCVLTETKORFPNNDKFKKLTITIDYFKNKRYFLERLENDTKEP 395

Qy 363 IQVRVDEDPGRFOHVMQDISSCTQKELRPEVDLVERSRHV--NTPKHAVALVTSLNA 420

Db 396 V-----DTQKCNIEHIMQTLTLPKQBDLGENFOAIHEKYIHTIGN-----LTLTGYSN 444

Qy 421 GYAEV-----LKS-----YMEYPTST 437

Db 445 KYSNNSFOEKHMEKGFQSSILKLNLSKLESFGEKEIEKRASDLADWALKIMYPTILE 504

Qy 438 GEIGVHOPSOEQYOQTEKKMHNGKALAEWLLS 471

Db 505 AETLEBYKPKKKEKKKEEYKIKKEKKYVLDLS 538

RESULT 14

335617

probable origin-binding protein - African swine fever virus

C:Species: African swine fever virus, ASFV

C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 08-Oct-1999

C:Accession: S35617

R:Sussman, M.D.; Lu, Z.; Kutish, G.F.; Afonso, C.A.; Rock, D.L.

Nucleic Acids Res. 21, 2254, 1993

A:Title: The identification of an African swine fever gene with conserved hellicase mo

A:Reference number: S35617; MUID:93275762; PMID:8389041

A:Accession: S35617

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1102 <SUS>

A:Cross-references: EMBL:L12174; NID:9289178; PIDN:AAA42719.1; PID:9289179

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1993

Query Match 3.7%; Score 110; DB 2; Length 1102;

Best Local Similarity 19.0%; Pred. No. 2.7;

Matches 103; Conservative 94; Mismatches 210; Indels 136; Gaps 26;

Qy 15 TTGGSKSVNFSLELD-MKYSSGTMKLTFRFTTCLIFSVLVAF-----SMIFQNH 65

Db 251 TTSSSCDSIQOSELEVLKFA--AGTLCKNHCF--LRVYKNLVLEKRPSPYCEICKRMH 305

Qy 66 PSDNRINGFAFARVLDAAGVFPVNTINISDKILGGLA--SGFDE---DCLISR---YQ 116

Db 306 DKDNTLLI---RVTKGRKRYQCHRDHDKHSILMCSLSTGNNFVETVEYQVMSKIEVHE 360

Qy 117 SVHYKRPSPYKPSYLSIKLRNYE-----KLHRCGPGTESYKALKQDQEHIDGPG 169

Db 361 SLTFEELPDQKHLYDESSMRTEYEVPTLVYKAQMKIGKTIO---LRNYLQKTYGDS 415

Qy 170 ECKY--VVMISFGNRLS--LASVELYALLT-----DRVLLYDRGKMDLFCERPF 219

Db 416 ISKQQTIFVFTRQJFSKNIGQRLPNFTLYSEVGTDLDSYERVIL-----QVESL----F 466

Qy 220 LGMSLLLDLDPMTQDFQGL-NQESSRCYGVW-----VKNOYIDTEGTLNLY 266

Db 467 RLTSTAEPVLDLILDEVESIFNQFNSGLHKYFAPSALEMMLETAHVICLADAMVKNRT 526

Qy 267 LHLVNDY-GD-----HDKMFCEGDDTFFGKVPMLYKTDNYFVPSLMLIP 311

Db 527 YNIIORFGVDVPIFFHMQYOQAQNDMYFTSSRREIMNLNKLKDLDEKKIYIPNSIME 586

Qy 312 G--FDEDLNKLFPQKATVPH-----HLGRYLFHPTNQVMGLV-----TRYEAVLSHAD 358

Db 587 ARLLQTFIQKRFPEKIKFYSSKSAHERESHFNWYWGILIDILYTPITSAGVYED 646

Qy 359 EKIGIQAVPDEDPGRFOHVMQDISSCTQ-----KEKILPEVDLV 399

Db 647 KR-----FDVLGYEFNMSCDVEPTCCOMLGRVRLSKCYKICLOGKONYFPET---I 696

Qy 400 ERSRHVNTPKHKAVALVTSINAGYAENLKSMEWYPTSTGEILIGVHQPSOEGYOQTEKKMH 459

Db 697 EDIEMFTLQKKNDTLFQITISN-----HQLSFYCKEKGRTPIYTYTHLMLETRIOH 749

Qy 460 NGK 462

Db 750 LSK 752

RESULT 15

C82183

catalse VC1585 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: C82183

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: C82183

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-567 <HEI>

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:29:59 ; Search time 31 Seconds
(without alignments)
529.613 Million cell updates/sec

Title: US-10-037-311a-1
Perfect score: 3004
Sequence: 1 MDNSYRRSSPIRTTGG5.....GIVPHVHCEDISWGLKLV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	3.1	723	4 US-09-134-001C-5060	Sequence 5060, Ap
2	92.5	3.1	993	4 US-08-836-667B-30	Sequence 30, Appl
3	92	3.1	809	4 US-09-186-276B-58	Sequence 58, Appl
4	92	3.1	809	4 US-08-842-445-58	Sequence 58, Appl
5	92	3.1	809	4 US-09-186-188B-58	Sequence 58, Appl
6	90	3.0	758	2 US-08-222-617A-6	Sequence 6, Appl
7	90	3.0	3666	2 US-08-222-617A-12	Sequence 12, Appl
8	90	3.0	3727	2 US-08-222-617A-27	Sequence 27, Appl
9	90	3.0	3778	4 US-08-222-617A-2	Sequence 2, Appl
10	89.5	3.0	1285	2 US-09-268-140-3	Sequence 3, Appl
11	89	3.0	1285	2 US-08-540-406-6	Sequence 6, Appl
12	89	3.0	1285	2 US-08-656-055-6	Sequence 6, Appl
13	89	3.0	1285	4 US-08-954-668-6	Sequence 6, Appl
14	89	3.0	1285	4 US-08-918-658-6	Sequence 6, Appl
15	89	3.0	1285	5 PCT-US95-13233-6	Sequence 6, Appl
16	88.5	2.9	515	3 US-08-369-822C-23	Sequence 23, Appl
17	88.5	2.9	515	3 US-08-582-776C-38	Sequence 38, Appl
18	88.5	2.9	515	3 US-08-434-831B-35	Sequence 35, Appl
19	86.5	2.9	771	4 US-09-462-284-2	Sequence 35, Appl
20	85	2.8	461	4 US-09-134-001C-5311	Sequence 5311, Ap
21	85	2.8	584	1 US-08-179-738-7	Sequence 7, Appl
22	85	2.8	584	1 US-08-628-145-7	Sequence 7, Appl
23	85	2.8	591	1 US-08-179-738-5	Sequence 5, Appl
24	85	2.8	591	1 US-08-628-145-5	Sequence 5, Appl
25	85	2.8	596	1 US-08-179-738-2	Sequence 2, Appl
26	85	2.8	596	1 US-08-628-145-2	Sequence 2, Appl
27	85	2.8	777	2 US-08-874-678-3	Sequence 3, Appl

28	85	2.8	777	3 US-08-643-839-3	Sequence 3, Appl
29	85	2.8	777	4 US-09-348-886-3	Sequence 3, Appl
30	85	2.8	1298	1 US-08-222-616-33	Sequence 33, Appl
31	85	2.8	1298	1 US-08-340-011-2	Sequence 2, Appl
32	85	2.8	1298	4 US-08-901-710-2	Sequence 2, Appl
33	85	2.8	1298	4 US-08-446-648-33	Sequence 33, Appl
34	85	2.8	1298	5 PCT-US95-04228-33	Sequence 33, Appl
35	85	2.8	1363	1 US-08-340-011-4	Sequence 4, Appl
36	85	2.8	1363	2 US-08-874-678-32	Sequence 32, Appl
37	85	2.8	1363	3 US-08-643-839-32	Sequence 32, Appl
38	85	2.8	1363	3 US-08-901-710-4	Sequence 4, Appl
39	85	2.8	1363	4 US-09-348-886-32	Sequence 32, Appl
40	84.5	2.8	848	1 US-08-045-806-4	Sequence 4, Appl
41	84.5	2.8	848	1 US-08-366-051B-4	Sequence 4, Appl
42	84	2.8	591	1 US-08-179-738-10	Sequence 10, Appl
43	84	2.8	591	2 US-08-628-145-10	Sequence 10, Appl
44	84	2.8	595	1 US-08-171-718-16	Sequence 16, Appl
45	84	2.8	595	3 US-08-478-087-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-5060
; Sequence 5060, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5060
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5060

Query Match 3.1%; Score 93; DB 4; Length 723;
Best Local Similarity 20.9%; Pred. No. 1.1;
Matches 110; Conservative 60; Mismatches 168; Indels 188; Gaps 28;

QY	113	SRVQSVHYRRKPPKPPSSYLIS----	KLBNYEKLHRCGPTESYKKALKOLDOE--	HTD 166	
DB	212	SHODEINTFKPEKTYTTLSTINVDGYDKL--	NOOKRYKD-----	KKLELLEHETKHQOE 262	
QY	167	GDGECKYVWISFSGIGNRILSLASVFLATLTD	RVLLVDRGKDDMLFCEPFLGMSWLL	226	
DB	263	G-----	KILEV-----	KKKNK-----	KSVNQ 278
QY	227	PLDFPMTDFDGLNDESSKCYWYKNOYIDTE	GLSHLY--LALVHDGDHMKFCESGD	285	
DB	279	PL-FMLTD-----	LOODEAVKRYKMGPK-----	ETLNLLOHLYERHKLVTYPRPTSNYLTDDM	329
QY	286	QFFICKVPMVLIYKTNYFVPSLMLPGFDE	LNTKLFPOKATVF-----	HH-----	330
DB	330	VDTIOERLAILATD-----	YKSHVADLISKSFSKMHIFNNOKYSDHAIITPEVR	381	
QY	331	-----	LGRYLFHPTNQVGLVTRY--YEAYLSHAD--	-----	EKIGIQVFEDE 370
DB	382	PSIEDLSQREFKTYMLIERPLENLMNPYLEV	LVLTTHAQKDYHFLVLEKKIPQL-----	436	
QY	371	DGPRQHWMDIISCTQ-----	KEKLLPEYDVLVENSRRHYNPKH--	KAVLYTSL-----	418
DB	437	---GYKALKDOLSHTLTHSFKEGOLFVHR	ITEIHETKAPDEYFNESGLKAMENPONH	493	

```

OY 419 ---NAGAAENIKSMWYEPYSTGIIIV-----HOPSOEYQOQTEKMHNGKA 463
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 494 IDLNKRYAKTLK-----HTGGIGVATRADILIEKIFMNALESRODKIYTSNGQ 545
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 464 LAEWYLITLONLWSM-SFPGYVAGGLGLKRWI-----LYRPE 503
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 546 ILESPELSTSPILTAONEEKLMIIEKKVNSQAFIDEMKNFFPKVYVNRKTSSEQKYHD 605
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 504 NRTTFD-PSCGRAM-----SMEPCSHPEFYDCAKGTGIDTGT 540
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 606 NLTTECEPTCGFMIRKYVTKNGOALVC-QDD--KCKTKNRIOKRT 647
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

```

1 RESULT 2
2 US-08-836-687B-30
3 : Sequence 30, Application US/08836687B
4 : Patent No. 6448034
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Gasson, Michael John
9 :
10 : APPLICANT: Dodd, Helen Mair
11 :
12 : TITLE OF INVENTION: PRODUCTION OF VARIANT NISIN
13 :
14 : FILE REFERENCE: 20147/7/0
15 :
16 : CURRENT APPLICATION NUMBER: US/08/836,687B
17 :
18 : CURRENT FILING DATE: 1995-11-20
19 :
20 : NUMBER OF SEQ ID NOS: 51
21 :
22 : SOFTWARE: Patentin Ver. 2.1
23 :
24 : SEQ ID NO 30
25 :
26 : LENGTH: 993
27 :
28 : TYPE: PRT
29 :
30 : ORGANISM: Lactococcus sp.
31 :
32 : US-08-836-687B-30

```

Query Match	3.1%;	Score 92.5;	DB 4;	Length 993;
Best Local Similarity	10.1%;	Score 92.5;	DB 4;	Length 993;

Local similarity	13.19;	Prod. NO. 2.1;							
Matches	90;	Conservative	63;	Mismatches	152;	Indels	167;	Gaps	23

```

OY      89  VTINISOKLIGLILASGPDDESCSLRQSVHYKRP--SPKPSBYSLIKIRNEYKHLKRC 146
Db      224  ISNLOKD-----LISDESNMFTFLTKVEALDEKDKYIIPKKOKQKFOESELE-----I 272
OY      147  GPGTESYKKALKODQ-----EHIDGDECKYVWVTSFGSLGRILSLAS 191
Db      273  GEGIEKLEKIEQVMSQILENDNYIQIDLLISDELINFDVQKQOOLEHAELEAGTYSVR 332
OY      192  VFL--YALITPRVLL---VDRKMDLDFEERPLGMSWLLPDPYKTDQFDG-----LN 240
Db      333  TYLDY---KDKFEKYGVDQEOVALDELFPDSTF-GIGAPARYNHPRNDPESESPSLAYS 388
OY      241  QESSRCGYM-----VKN-QVLDTEGTLSH-----LYLHVIHDYGHDKM 279
Db      389  EERSEKTLMSVEAVKNNHNYINDDLESHQOKDLKESLSQLELFINIAKEY---EKD 445
OY      280  FFCGSDOTFIGKVPWLLIVKTDNYFVBSLMLPGFDELNKLKPFOKATVENHILGRYLFHPT 339
Db      446  IFILGD-----IVGNNN-----LGSAGRSFSLASPE----- 471
OY      340  NOVWGLVTRYRYEAVLISHADEKIGIQRVFPEDGPPRQHWADQSSCTOKREKLLPEVDYLV 399
Db      472  ---LTSIHRTIVDSVER-----ENENKELTSSC-----ELVFPLP 501
OY      400  ERSRHHV-----TPKHRAVLVTSLNAGYAENLKSWMYWEPTSGEIIIG 443
Db      502  ENIHAAVWMTSILRRKVLPEFTSTSHNEVLNENIYIGIDE--KEKFYARODSTQEVLEKF 559
OY      444  HQPQEQEYQOTEEKKMHN---GKALAEWYLLSLTDWLVTSAMS---TPGYV 487
Db      560  YIYS-----MYNKLFSNELLEFLEISLDDKFGNLPWELLRYDRDYI 601

```

RESULT 3
US-09-186-276B-58
; Sequence 58, Application US/09186276B
; Patent No. 6388173

```

1  ? GENERAL INFORMATION:
2  ? APPLICANT: Benfey, Philip
3  ? APPLICANT: Dilaurenzio, Laura
4  ? APPLICANT: Wyszocka-Diller, Joanna
5  ? APPLICANT: Malamy, Jocelyn E.
6  ? APPLICANT: Pysht, Leonard
7  ? APPLICANT: Helarintula, Yrjo
8  ? TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
9  ? FILE REFERENCE: 5914-075-999
10 ? CURRENT APPLICATION NUMBER: US/09/186,276B
11 ? CURRENT FILING DATE: 1998-11-05
12 ? PRIOR APPLICATION NUMBER: 08/842,445
13 ? PRIOR FILING DATE: 1997-04-24
14 ? PRIOR APPLICATION NUMBER: 08/638,617
15 ? PRIOR FILING DATE: 1996-04-26
16 ? NUMBER OF SEQ ID NOS: 79
17 ? SOFTWARE: PastSeq for Windows Version 3.0
18 ? SEQ ID NO 58
19 ? LENGTH: 809
20 ? TYPE: PRT
21 ? ORGANISM: Arabidopsis thaliana
22 ? FEATURE:
23 ? NAME/KEY: VARIANT
24 ? LOCATION: (1)...(809)
25 ? OTHER INFORMATION: Xaa = Any Amino Acid
26 US-09-186-276B-58

```

Query Match	3.18;	Score 92;	DB 4;	Length 809
Best local similarity	21.79;	Pred No	17.	

2006 Local Unimilarity 21.7%, Freq. NO. 1.7/
Matches 84; Conservative 52; Mismatches 179; Indels 72; Gaps 18,

```

QY 66 PDSRIRRMFGFEARLD-----AGVFP---NTJNMSDXLGLGLASSGFE-----108
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 63 PSSNQTLGTLANGCYLDDLDSSLDPEARYPSQNNNNNNNNKKAIVAGDLSSSSDDFGS 122
    | : : : : : | : : : : : | : : : : : | : : : : : |
QY 109 DCLSRQSVHYRRKPSPYKPSYLIS- KLRNYEK-LHKRCGPTEYSKKAALKODOEHI- 165
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 DSVLKYISQVLMEDMEERKPCMFDALALQAEKSLYEALGEXKPPSSSASVYDPERLA 182
    | : : : : : | : : : : : | : : : : : | : : : : : |
QY 166 -DDGGECKYVWVWISFSLGRILSLASVFLYALLTRVLLVDRGKMDDLFCPEFLGMS 223
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 183 SHSPDGSCS-----GAFSDYAS-----TTTTSSHSWSDGLENRP-----S 221
    | : : : : : | : : : : : | : : : : : | : : : : : |
QY 224 WL--LPLDFEPTQDFGLJNDSRCYGVWYKNOVITEGLSHLYLHLVHDYDHDHXMFE 280
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 222 WLHPHPSNF-----VFQSTSRKSNVYTGCGGGGANSAYVGSFGDDLVSNNMKRD--DELAQ 275
    | : : : : : | : : : : : | : : : : : | : : : : : |
QY 281 FCEGDO--FTIGKVPMLVKTQDVFYFPSLMLTPGFDEINKLP---PQKATVFNNHGRY 334
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 276 FKKEVEASKRLPRSSQLFDVDSY-IP-----MNGSKSEKNSSEVYVKTEKKEDDEYHHNHS 331
    | : : : : : | : : : : : | : : : : : | : : : : : |
QY 335 LEHPTNOVWGLVTRYAYVLYSHADEKTIGIOVRVFEDEDPGFQHVMDIISCTQKEKLLPE 394
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 332 YAPPNNRLTKGKSHMRDEDEDFEERSNKOSAYVEE--SELSEWFDMFLCGPGKPYC-- 388
    | : : : : : | : : : : : | : : : : : | : : : : : |
QY 395 VDLIVERSRHVNPCKHKAIVLTLINAG 421
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 389 -----ILNQNFETSAKAVTAOSNG 408
    | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

RESULT 4
US-08-842-445-58
Sequence 58, Application US/08842445A
Patent No. 6441270
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scatterbow Gene, Promoter and Uses
FILE REFERENCE: 5914-056-999
CURRENT APPLICATION NUMBER: US/08/842,445A
CURRENT FILING DATE: 1997-04-24
EARLIER APPLICATION NUMBER: 08/638,617
EARLIER FILING DATE: 1996-04-26

```


NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 58
LENGTH: 809
TYPE: PRF
ORGANISM: Plant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(809)
OTHER INFORMATION: xaa = Any Amino Acid
US-08-842-445-58

Query Match 3.1%; Score 92; DB 4; Length 809;
Best Local Similarity 21.7%; Pred. No. 1.7;
Matches 84; Conservative 52; Mismatches 179; Indels 72; Gaps 18;

QY 66 PSDSNRIMGFARVLD-----AGVFP-----NTNINSKLLGGLASGDE---- 108
DB 63 PPSSNQTGLANGFYLDLDFSSLDPEAYPSQNNNNNNINNKAVAGDLLSSSSDDADFS 122
QY 109 DSCLSRQSVHYRKPSPYKSSYLIS-KLRNYEK-LHKRCGPTESYKKALKQDOEHI- 165
DB 123 DSVLKYSQVLMEDMEKPCMFHDALALQAEKSLYEALGEKDPSSSSASVDHPERLA 182
QY 166 --DGDGCKYVWVISFSGLGNRLSLASVFLYALITDVRLLVDRGKMDDLFCPEFLGMS 223
DB 183 SHSPDSCS-----GGAFSDYAS-----TTTTSSDHSWVDGLENRP-----S 221
QY 224 WL---LPDLPMTDQFGLNOESSRCYGYWVKNOVIDTEGTLSHLYLHLVHDYGDHDKMF 280
DB 222 WLHTPMSNF---VFQSTRSNSVTGGGGGNSAVYGSFGDDLVSNMFKD--DELAHQ 275
QY 281 FCEGDO---FPIGVPMILVKTQNYFVPSLMLIPGPDDELKLP---PQKATVFHHLGRY 334
DB 276 FKQVGEASRFLPKSQLFIDVDST-IP---MSGSKENGSEVVFVTEKEDTEHHHHS 331
QY 335 LFHPTNQWGLVTRYEAYLSHADEKIGIQRVDEDPGPFQHYMDQISSCTQKEKLP 394
DB 332 YAPPNNLTGKKSHMRDEDEDFEERSNKOSAVYEE-SELSEMFDMFLCGPKRPVC-- 388
QY 395 VDTLVERSRYVTPKHAVALVTSINAG 421
DB 389 -----ILNQNFPTESAKVYTAQSNQ 408

RESULT 5
US-09-186-188B-58
Sequence 58, Application US/09186188B
Patent No. 6455672
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
FILE REFERENCE: 5914-074-999
CURRENT APPLICATION NUMBER: US/09/186,188B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 58
LENGTH: 809
TYPE: PRF
ORGANISM: Plant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(809)
OTHER INFORMATION: xaa = Any Amino Acid
US-09-186-188B-58

Query Match 3.1%; Score 92; DB 4; Length 809;

Best Local Similarity 21.7%; Pred. No. 1.7;
Matches 84; Conservative 52; Mismatches 179; Indels 72; Gaps 18;

QY 66 PSDSNRIMGFARVLD-----AGVFP-----NTNINSKLLGGLASGDE---- 108
DB 63 PPSSNQTGLANGFYLDLDFSSLDPEAYPSQNNNNNNINNKAVAGDLLSSSSDDADFS 122
QY 109 DSCLSRQSVHYRKPSPYKSSYLIS-KLRNYEK-LHKRCGPTESYKKALKQDOEHI- 165
DB 123 DSVLKYSQVLMEDMEKPCMFHDALALQAEKSLYEALGEKDPSSSSASVDHPERLA 182
QY 166 --DGDGCKYVWVISFSGLGNRLSLASVFLYALITDVRLLVDRGKMDDLFCPEFLGMS 223
DB 183 SHSPDSCS-----GGAFSDYAS-----TTTTSSDHSWVDGLENRP-----S 221
QY 224 WL---LPDLPMTDQFGLNOESSRCYGYWVKNOVIDTEGTLSHLYLHLVHDYGDHDKMF 280
DB 222 WLHTPMSNF---VFQSTRSNSVTGGGGGNSAVYGSFGDDLVSNMFKD--DELAHQ 275
QY 281 FCEGDO---FPIGVPMILVKTQNYFVPSLMLIPGPDDELKLP---PQKATVFHHLGRY 334
DB 276 FKQVGEASRFLPKSQLFIDVDST-IP---MSGSKENGSEVVFVTEKEDTEHHHHS 331
QY 335 LFHPTNQWGLVTRYEAYLSHADEKIGIQRVDEDPGPFQHYMDQISSCTQKEKLP 394
DB 332 YAPPNNLTGKKSHMRDEDEDFEERSNKOSAVYEE-SELSEMFDMFLCGPKRPVC-- 388
QY 395 VDTLVERSRYVTPKHAVALVTSINAG 421
DB 389 -----ILNQNFPTESAKVYTAQSNQ 408

RESULT 6
US-08-222-617A-6
Sequence 6, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 758 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: ORGANISM: Penicillium chrysogenum
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 1..758
: OTHER INFORMATION: /label= Domain II
: OTHER INFORMATION: /note= "Domain II of ACV Synthetase from
: OTHER INFORMATION: Penicillium chrysogenum; aa 1397-2154"
:
: US-08-222-617A-6
:
Query Match
Best Local Similarity 20.7%; Score 90; DB 2; Length 758;
Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps 13;

QY 71 RIMGFAERVLADAGVFPWNTNINSKLLGLLGSFDEDSCLSRYSVHYRKPSPFPKS- 129
Db 132 RIKGMAASGTL---LPSVLPANPD-----SKWS---VSNPSPLSRST 168
QY 130 --SYLISKRLRYEKLHKRCGPGTESYKALKQLDQEHIDSGECKYVYVWISFSGIKRIL 187
Db 169 DLAYIIIT-----SGTTGRPKG-----VTYEHHRGVNLDQV 198
QY 188 SLASVFLYALLTDRLVLDVRGKMD---DLFCEPFLGMSWLLPLDFPMTDQFDGLNDSS 244
Db 199 SLKSVFGLRDTDEVIILFSNRYVDFHFEQMTDALNGOTLLVNL-----DGMRGDKE 251
QY 245 RCYGYWAKNOYIDTEGTLSHLYLHLVHDYGDHDKMFCEGD-----QFTIGKVP 293
Db 252 RLRYRIEKNRYTLSDGTPSVSMYEFSEFKDHLRRVDCVGEAFSEPVDKIRETFHGLV- 310
QY 294 WLIVKTDNVEFVPSLMLIPGFDELNKLFPKATVFNHILGRFLPHPTNOVGLVTRYEAY 353
Db 311 -----INGYGPTEVSTITTKRLYPF-PERRM----- 335
QY 354 LSHADEKIGIQVR-----VFDED 371
Db 336 ----DKSIGQGVHNSTSYLVNED 354

RESULT 7
US-08-222-617A-12
: Sequence 12, Application US/08222617A
: Patent No. 5882879
: GENERAL INFORMATION:
: APPLICANT: Veenstra, Annemarie E.
: APPLICANT: Martin, Juan F.
: APPLICANT: Garcia, Bruno D.
: APPLICANT: Gutierrez, Santiago
: APPLICANT: Barredo, Jose L.
: APPLICANT: Von Doehren, Hans
: APPLICANT: Palissa, Harriet
: APPLICANT: Van Liempt, Henk
: APPLICANT: Montenegro, Eduardo P.
: TITLE OF INVENTION: A Method for Influencing Beta-Lactam
: TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
: TITLE OF INVENTION: Quantities of ACV Synthetase
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
: STREET: 300 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
```

```

: APPLICATION NUMBER: US/08/222,617A
: FILING DATE: 04-APR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: REFERENCE/DOCKET NUMBER: 97,157
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3666 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Penicillium chrysogenum
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..3666
: OTHER INFORMATION: /label= region
: OTHER INFORMATION: /note= "Region of ACV Synthetase; aa 62-3727"
:
: US-08-222-617A-12
:
Query Match
Best Local Similarity 20.7%; Score 90; DB 2; Length 3666;
Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps 13;

QY 71 RIMGFAERVLADAGVFPWNTNINSKLLGLLGSFDEDSCLSRYSVHYRKPSPFPKS- 129
Db 1467 RIKGMAASGTL---LPSVLPANPD-----SKWS---VSNPSPLSRST 1503
QY 130 --SYLISKRLRYEKLHKRCGPGTESYKALKQLDQEHIDSGECKYVYVWISFSGIKRIL 187
Db 1504 DLAYIIIT-----SGTTGRPKG-----VTYEHHRGVNLDQV 1533
QY 188 SLASVFLYALLTDRLVLDVRGKMD---DLFCEPFLGMSWLLPLDFPMTDQFDGLNDSS 244
Db 1534 SLKSVFGLRDTDEVIILFSNRYVDFHFEQMTDALNGOTLLVNL-----DGMRGDKE 1586
QY 245 RCYGYWAKNOYIDTEGTLSHLYLHLVHDYGDHDKMFCEGD-----QFTIGKVP 293
Db 1587 RLRYRIEKNRYTLSDGTPSVSMYEFSEFKDHLRRVDCVGEAFSEPVDKIRETFHGLV- 1645
QY 294 WLIVKTDNVEFVPSLMLIPGFDELNKLFPKATVFNHILGRFLPHPTNOVGLVTRYEAY 353
Db 1646 -----INGYGPTEVSTITTKRLYPF-PERRM----- 1670
QY 354 LSHADEKIGIQVR-----VFDED 371
Db 1671 ----DKSIGQGVHNSTSYLVNED 1689

RESULT 8
US-08-222-617A-27
: Sequence 27, Application US/08222617A
: Patent No. 5882879
: GENERAL INFORMATION:
: APPLICANT: Veenstra, Annemarie E.
: APPLICANT: Martin, Juan F.
: APPLICANT: Garcia, Bruno D.
: APPLICANT: Gutierrez, Santiago
: APPLICANT: Barredo, Jose L.
: APPLICANT: Von Doehren, Hans
: APPLICANT: Palissa, Harriet
: APPLICANT: Van Liempt, Henk
: APPLICANT: Montenegro, Eduardo P.
: TITLE OF INVENTION: A Method for Influencing Beta-Lactam
: TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
: TITLE OF INVENTION: Quantities of ACV Synthetase
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
: STREET: 300 South Wacker Drive
```

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3727 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-27

Query Match 3.0%; Score 90; DB 2; Length 3727;
Best Local Similarity 20.7%; Pred. No. 34;
Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps 13;

OY 71 RMGFAEARVLDAGVFPPVNTNNSDKLLAGLLASGFDESDCLSRQSVHYRKPSPYKPS- 129
DB 1528 RIKGMAASGTL---LPSVLPAHPD-----SKMS---VSNPSPLSRST 1564
OY 130 --SYLISKLRNVEKLRHKGCGTESYKRALQDQEHIDGDECKYVWVMSFGGLGNRL 187
DB 1565 DLATIIYF-----SQTIGRPKG-----YVEIHGAVNNLOV 1594
OY 188 SLASVFLALLTDRLVLDVDRGKMD--DLCEPFLGSMWLLPLDFPMTDOFGDNGESS 244
DB 1595 SLSKVFGLMDTDEVLILFSNVVDFHFVEQMTDALINCGTLLVLN-----DKMRGDK 1647
OY 245 RCYGVNKNQVYIDTGGTSLHLVHLVHDYGDHDKMFEGCD-----QTFIGKVP 293
DB 1648 RLYRYIEKRNRYVLSGTPSVVSMYFSPRKHDLRVDCVGEAFSESPVDEKIRETHGLV- 1706
OY 294 WLIVKTDNVFVPSLMLDGFDELKMLFPQKATVFHNLGRYLFHPTNQVWGLVTRYEAY 353
DB 1707 -----INGYGPTEVSITTHKRLYFF-PERRR----- 1731
OY 354 LSHADEKIGIOVR-----VFDED 371
DB 1732 ----DKSIGQGVHNSTSYLVNED 1750

RESULT 9
US-08-222-617A-2
Sequence 2, Application US/08222617A
Patent No. 5882879

GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A method for influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of Acv Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3778 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-2

Query Match 3.0%; Score 90; DB 2; Length 3778;
Best Local Similarity 20.7%; Pred. No. 35;
Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps 13;

OY 71 RMGFAEARVLDAGVFPPVNTNNSDKLLAGLLASGFDESDCLSRQSVHYRKPSPYKPS- 129
DB 1528 RIKGMAASGTL---LPSVLPAHPD-----SKMS---VSNPSPLSRST 1564
OY 130 --SYLISKLRNVEKLRHKGCGTESYKRALQDQEHIDGDECKYVWVMSFGGLGNRL 187
DB 1565 DLATIIYF-----SQTIGRPKG-----YVEIHGAVNNLOV 1594
OY 188 SLASVFLALLTDRLVLDVDRGKMD--DLCEPFLGSMWLLPLDFPMTDOFGDNGESS 244
DB 1595 SLSKVFGLMDTDEVLILFSNVVDFHFVEQMTDALINCGTLLVLN-----DKMRGDK 1647
OY 245 RCYGVNKNQVYIDTGGTSLHLVHLVHDYGDHDKMFEGCD-----QTFIGKVP 293
DB 1648 RLYRYIEKRNRYVLSGTPSVVSMYFSPRKHDLRVDCVGEAFSESPVDEKIRETHGLV- 1706
OY 294 WLIVKTDNVFVPSLMLDGFDELKMLFPQKATVFHNLGRYLFHPTNQVWGLVTRYEAY 353
DB 1707 -----INGYGPTEVSITTHKRLYFF-PERRR----- 1731
OY 354 LSHADEKIGIOVR-----VFDED 371
DB 1732 ----DKSIGQGVHNSTSYLVNED 1750

RESULT 10
US-09-268-140-3
Sequence 3, Application US/09268140
Patent No. 6268176

GENERAL INFORMATION:
APPLICANT: Gemmill, Robert M.
APPLICANT: Drabkin, Harry A.
TITLE OF INVENTION: TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED
FILE REFERENCE: 93445-00004
CURRENT APPLICATION NUMBER: US/09/268,140
CURRENT FILING DATE: 2000-03-12
PRIOR APPLICATION NUMBER: US 60/077,723
PRIOR FILING DATE: 1998-03-12
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1286
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-268-140-3

Query Match 3.0%; Score 89.5; DB 4; Length 1286;
Best Local Similarity 20.6%; Pred. No. 6.8;
Matches 64; Conservative 41; Mismatches 81; Indels 125; Gaps 17;

QY 218 PRLGMSW-----LRLDPRMTDQFDGLN-----QESSRCGYM 250
||| ||| : : : ||| : : : |||
DB 672 PRLMSWVKFELTVMGFLAALLISLVASTRLODGLDIDLVPKDSNEHKFLDAQTRLFGEY 731
QY 251 VKNQYIDEGTSLH-LYLHLVDHDXGDHDKMFCFEGDQTFIGVPMILVKTBNYFVPSLML 309
||| ||| : : : ||| : : : |||
DB 732 SMAYV--TQGNFEYTPQOQLNDY--HDSF-----RVPH-VIKNDNGGLPDEFWL 776
||| ||| : : : ||| : : : |||
QY 310 IEGFDELKLEPKATVFNHIGRYLFHPTNQWGLVTRYEAYLSHADEKIGIQVRED 369
||| ||| : : : ||| : : : |||
DB 777 L-----LF-----SEWLGNIQKIFD 791
QY 370 EDGPGFQHVMDQISSCTQKEKLLPEVDI-----LVRSRVNTPKRAVLVTS--LNA 420
||| ||| : : : ||| : : : |||
DB 792 E-----EYRDGRLEKCFPNASSDALAYKLIQVOTGHVDPVDEKELVITNRLVNS 842
QY 421 GVAENLKSMY-----WEYPTSTGEITGVHQPQSEGVOQTEKKMHNGKAL-----A 465
||| ||| : : : ||| : : : |||
DB 842 DGTINORAFYNTLSAMATNDVFAYGASQGL--YEPROYFHQPNEDYDLKIPKSLPLVYA 900
QY 466 EM--YLLSLTD 474
||| ||| |||
DB 901 QMPFYHLGLTD 911

RESULT 11
US-08-540-406-6
; Sequence 6, Application US/08540406
; Patent No. 5837538
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540.406
; FILING DATE: 06-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertam I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1285 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-540-406-6

Query Match 3.0%; Score 89; DB 2; Length 1285;
Best Local Similarity 20.3%; Pred. No. 7.6;

Matches 63; Conservative 39; Mismatches 83; Indels 126; Gaps 16;

QY 218 PRLGMSW-----LRLDPRMTDQFDGLN-----QESSRCGYM 250
||| ||| : : : ||| : : : |||
DB 672 PRLMSWVKFELTVMGFLAALLISLVASTRLODGLDIDLVPKDSNEHKFLDAQTRLFGEY 731
QY 251 VKNQYIDEGTSLH-LYLHLVDHDXGDHDKMFCFEGDQTFIGVPMILVKTBNYFVPSLML 309
||| ||| : : : ||| : : : |||
DB 732 SMAYV--TQGNFEYTPQOQLNDY--HDSF-----RVPH-VIKNDNGGLPDEFWL 775
||| ||| : : : ||| : : : |||
QY 310 IEGFDELKLEPKATVFNHIGRYLFHPTNQWGLVTRYEAYLSHADEKIGIQVRED 369
||| ||| : : : ||| : : : |||
DB 776 L-----LF-----SEWLGNIQKIFD 790
QY 370 EDGPGFQHVMDQISSCTQKEKLLPEVDI-----LVRSRVNTPKRAVLVTS--LNA 420
||| ||| : : : ||| : : : |||
DB 791 E-----EYRDGRLEKCFPNASSDALAYKLIQVOTGHVDPVDEKELVITNRLVNS 841
QY 421 GVAENLKSMY-----WEYPTSTGEITGVHQPQSEGVOQTEKKMHNGKAL-----A 465
||| ||| : : : ||| : : : |||
DB 842 DGTINORAFYNTLSAMATNDVFAYGASQGL--YEPROYFHQPNEDYDLKIPKSLPLVYA 899
QY 466 EM--YLLSLTD 474
||| ||| |||
DB 900 QMPFYHLGLTD 910

RESULT 12
US-08-656-055-6
; Sequence 6, Application US/08656055
; Patent No. 6027882
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/656.055
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/540.406
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertam I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1285 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-656-055-6

Query Match 3.0%; Score 89; DB 3; Length 1285;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:28:54 ; Search time 38 Seconds
(without alignments)
3025.639 Million cell updates/sec

Title: US-10-037-311a-1
Perfect score: 3004
Sequence: 1 MDONSYYRRSSSPIRTTGGG.....GTLVPHVRHCEDISWGKLKV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	279.5	9.3	315	10	094LJ8
2	116.5	3.9	329	16	098AU6
3	115	3.8	370	2	09AMY4
4	114	3.8	1055	12	065146
5	112.5	3.7	2149	10	092W93
6	112.5	3.7	2244	12	0997F0
7	112.5	3.7	2244	12	0914E5
8	110.5	3.7	322	2	085712
9	110	3.7	529	5	021310
10	110	3.7	658	16	025252
11	110	3.7	1102	12	007183
12	107	3.6	567	16	09KRO1
13	106.5	3.5	80	10	09SJP5
14	106	3.5	3844	5	094648
15	105	3.5	381	5	001660
16	104.5	3.5	898	3	09C213

17	104.5	3.5	1583	5	076222	076222 trypanosoma
18	103.5	3.4	1209	5	021667	021667 caenorhabdi
19	103	3.4	611	13	091BF6	091BF6 xenopus lae
20	102.5	3.4	382	12	063594	063594 bovine coro
21	102.5	3.4	413	2	093PC6	093PC6 microscilla
22	102	3.4	395	4	09BSW2	09BSW2 homo sapien
23	102	3.4	611	13	09XOV8	09XOV8 thermotoga
24	101	3.4	811	16	09PTI0	09PTI0 xenopus lae
25	101	3.4	823	5	08ZPJ7	08ZPJ7 salmonella
26	101	3.4	823	5	09USP4	09USP4 naemonchus
27	100.5	3.3	409	16	08X5U5	08X5U5 escherichia
28	100.5	3.3	475	2	09X564	09X564 enterococu
29	100.5	3.3	599	16	08XKE7	08XKE7 clostridium
30	99.5	3.3	439	5	09VBR5	09VBR5 drosophila
31	99	3.3	1502	10	09LNX6	09LNX6 arabidopsis
32	98.5	3.3	636	13	090Z16	090Z16 paracitnity
33	98.5	3.3	846	10	09C9H7	09C9H7 arabidopsis
34	98	3.3	824	5	045585	045585 caenorhabdi
35	98	3.3	864	4	09NVU3	09NVU3 homo sapien
36	98	3.3	1274	4	09P2D5	09P2D5 homo sapien
37	98	3.3	1383	4	09UHR6	09UHR6 homo sapien
38	97	3.2	515	4	09NWV7	09NWV7 homo sapien
39	97	3.2	678	16	09KBU7	09KBU7 bacillus ha
40	97	3.2	1163	4	09P2J6	09P2J6 homo sapien
41	97	3.2	1903	5	09U0N7	09U0N7 plasmodium
42	97	3.2	2044	16	08Y8Q4	08Y8Q4 listeria mo
43	97	3.2	4717	2	094Z48	094Z48 schizosacch
44	96.5	3.2	335	2	0916E0	0916E0 escherichia
45	96.5	3.2	768	12	09DR83	09DR83 bovine coro

ALIGNMENTS

RESULT 1	ID	094LJ8	PRELIMINARY:	PRT:	315 AA.
AC	094LJ8:				
DT	01-DEC-2001 (TREMBLrel. 19, Created)				
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Xyloglucan fucosyltransferase-like.				
OS	Oryza sativa (Rice).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:				
OC	Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae:				
OC	Ehharitoidae: Oryzae: Oryza.				
OX	NCBI_TaxID=4530;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Wing R.A., Fritsch D., Presting G., Wood T., Yu Y., Soderlund C.,				
RA	Kim H., Rambo T., Henry D., Simmons J.;				
RT	"Rice Genomic Sequence."				
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AC079128; AAK52525.1; -				
DR	InterPro: IPR004938; XG_Frase.				
DR	Pfam: PF03254; XG_Frase. 1.				
KW	Glycosyltransferase; Transferase.				
SO	SEQUENCE 315 AA; 34604 MW; 04AEDA5D735E965E CRC64;				
Query Match	9.3%;	Score 279.5;	DB 10;	Length 315;	
Best Local Similarity	33.2%;	Pred. No. 8.7e-16;			
Matches 68;	Conservative 26;	Mismatches 48;	Indels 63;	Gaps 6;	
OY	145	RCGPTESYKALKQDDEHIDGDECKYVWISFSGLNRLISLAVFLYALLTDVLL	204		
DB	29	RGAPETESYNRAVRLK---DGSKGSGATEADARCCSR-----ATSR---	68		
OY	205	VDRGDMDDLCCEPFLGSMWLLPDPMTDQFDGLNOSKRCYGMKNQYIDTGTSLH	264		
DB	69	-----WCRSYANPS-----ADSAESYGMKNKVLGTDGSGDD	101		
OY	265	L-----YLHLVHDYGDHDKMFCEGDDTFIGKVPWLIVKTDNYPVPSTLMLTPGDD	315		

Db 102 MEAQMFAFAVLAHLNHDYDDDKMEFCDDDR-----LVNKRTPIYVPSLIVTTFOD 154
 QY 316 ELNKLFPKATVFNHGLRYLPHPTN 340
 ||:||||: |||:|||||
 Db 155 ELDALEPBGAVFHYIGRLFFQAN 179

RESULT 2

Q98AU6 PRELIMINARY; PRT; 329 AA.
 AC 098AU6:
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Nodulation protein, NodZ.
 GN MR5848.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003007; BAB5226.1; -
 KW Complete proteome.
 SQ SEQUENCE 329 AA; 37449 MW; 2460E90207D38287 CRC64;

Query Match 3.9%; Score 116.5; DB 16; Length 329;
 Best Local Similarity 21.1%; Pred. No. 0.11;
 Matches 78; Conservative 44; Mismatches 123; Indels 125; Gaps 15;

QY 165 IDGDECKYVWISFGNRLSLASVFLALTDRLVLD-RGK-DMDDLFCPEPTGM 222
 Db 3 VDRSRNDRFVVSRRRTGLGDCLSASAMDYAHRTGRALADIMRGSCLIDNPFANAF--- 59
 QY 223 SWLPLDFPMTDQFGLNQESSRCGYWKNQVIDTEGTLSHLYLHVHDGDHDKMFEC 282
 Db 60 ---PAFKPIEDIGV-----VICDQINHV----- 83
 QY 283 EGDQTFICKVPMVLVKTNDYFVPSLMLPGFD-----DELNKLFF----- 321
 Db 84 ---SFGP-----FFPSWNNKPSIDCVYRPDEQIFQERDELNDLFGAQDCEA 128
 QY 322 -----POKATVFHNLGRYLFHPTNQVGLVTRYEATLSHAD-EKGIQYR 366
 Db 129 NTVVCDACLMRCDEKAE-----RTIFRSITPRVEIDREIALQEHFHDGSHIIGVHR 182
 QY 367 VEDEDPGFQHVMDQISSCTQKEKLLPEVDLVERSHVNTPKHKAIVTSLNAGYANL 426
 Db 183 -----HGNGEDIMDHPYVWADPKVALEQVCAIRNAKQPHSRPVRYLCLDSAKVVDHL 237
 QY 427 KSMWEYPTSGEITGIVQPSQEGYQOTEKMMN-----KALAEWYLLSLTDNLV- 477
 Db 238 SGVPELFTVPKRF-----QADQAG-----PLHSAALCTDGGSGALVEMYLLGRCDVYIR 287
 QY 478 ---TSAMSTF 484
 |||:|:
 Db 288 FPPTSATFTR 297

RESULT 3

Q9AMY4 PRELIMINARY; PRT; 370 AA.
 AC 09AMY4;
 ID 09AMY4;
 QY 09AMY4;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE NodZ.
 GN NODZ.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=110SPC4;
 RX MEDLINE=21101824; PubMed=11157954;
 RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,
 RA Hennecke H.,
 RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb
 RT DNA region of the Bradyrhizobium japonicum chromosome."
 RL J. Bacteriol. 183:1405-1412(2001).
 DR EMBL; AF322013; AAC61006.1; -
 SQ SEQUENCE 370 AA; 42144 MW; A76834433053C58E CRC64;

Query Match 3.8%; Score 115; DB 2; Length 370;
 Best Local Similarity 20.8%; Pred. No. 0.18;
 Matches 86; Conservative 56; Mismatches 144; Indels 128; Gaps 21;

QY 120 YRKPEPYKPSVLLS---KLRNYEKLKRCGPGTESYKALKQJLDQEHIDGDECKYVW 176
 Db 4 YRCSP-APRQAVTPGKRIKRETSVLSVOPGA-----REKARQWSSSNDREYVS 55
 QY 177 ISFSGLNRLSLASVFLYALTLDRVLVD-RGKMDDLFCPEPTGM 235
 Db 56 RRRGFGDCIWSLAAAMFAKQGTGLAIDWRG-----CY-----LDEPTNA 99
 QY 236 FDGLNQESSRCGYWKNQVIDTEGTLSHLYLHVHDGDHDKMFCEGDQTFICKVPM 295
 Db 100 FP-----VFEEVEDIG--GVNVICDD----- 120
 QY 296 IVKTDNY---FVPSLMLPGFD-----DELNKLFPKA-----TVF----- 328
 Db 121 -INRSFPGFPPTWNNKPSFDCIYRPDEQIFRERDLDQFQSORSDANTVVCDACLM 179
 QY 329 ---HNLGRYLFHPTNQVGLVTRYEATLSHAD-EKGIQYRVEDDPGFQHVMDQ 382
 Db 180 WRQGEAREIFRSIKRPELQARIDAIYREHFEYVIGIHVR-----HONGDINGHA 234
 QY 383 SSGCTQKEKLLPEVDLVERSHVNTPKH-KAVLVTSNAGYAEWYKSMWEYPTSGEIT 441
 Db 235 PYWADTERALROIYNAIDEARSLSHAKPVRAFLCTD-----SALVLEQVSVFP---DV 286
 QY 442 GVHDPQSGYQOTEKMMNG-----KALAEWYLLSLTDNLV---TSAMSTF 484
 Db 287 AI--PKQFQAPQAPLHHPALGAGGFSALTEMYLLAKCDTVIRPPTSATFTR 338

RESULT 4

Q65146 PRELIMINARY; PRT; 1055 AA.
 AC 065146;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Helicase.
 GN F1055L.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfarivirus.
 OX NCBI_TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=21820291; PubMed=11831707;
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinuela E.;
 RT "Analysis of the complete nucleotide sequence of African swine fever

RT virus.";
RL virology 208:249-278(1995).
[12]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94233765; PubMed=8178480;
RA De la Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;
RT "Nucleotide sequence and variability of the inverted terminal
RL repetitions of African swine fever virus DNA.";
RN virology 201:152-156(1994).
[13]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90219205; PubMed=2325203;
RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
RT De la Vega I., Blasco R., Vinuela E.;
RL "Multigene families in African swine fever virus: Family 360.";
RN J. Virol. 64:2073-2081(1990).
[14]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90219204; PubMed=2325202;
RA Almendral J.M., Almazan F., Blasco R., Vinuela E.;
RT "Multigene families in African swine fever virus: Family 110.";
RL J. Virol. 64:2064-2072(1990).
[15]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=91134988; PubMed=1994575;
RA Camacho A., Vinuela E.;
RT "Protein p32 of african swine fever virus: an early structural protein
RL that is incorporated into the membrane of infected cells.";
RN virology 181:251-257(1991).
[16]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX Almazan F., Murguia J.R., Rodriguez J.M., La Vega I., Vinuela E.;
RT "A set of African swine fever virus tandem repeats shares similarities
RL with SAR-like sequences.";
RN J. Gen. Virol. 0:0-0(0).
[17]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94187118; PubMed=8139051;
RA Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,
RT Vinuela E.;
RL "Multigene families in African swine fever virus: family 505.";
RN J. Virol. 68:2746-2751(1994).
[18]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93346971; PubMed=8339314;
RA Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
RT African swine fever virus thymidylate kinase gene: sequence and
RL transcriptional mapping.";
RN J. Gen. Virol. 74:1633-1638(1993).
[19]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94065656; PubMed=8245848;
RA Alcaml A., Angulo A., Vinuela E.;
RT "Mapping and sequence of the gene encoding the African swine fever
RL virion protein of M(r) 11500.";
RN J. Gen. Virol. 74:2317-2324(1993).
[10]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93277388; PubMed=8503790;
RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
RT "Structure and expression in E. coli of the gene coding for protein
RL p10 of African swine fever virus.";
RN Arch. Virol. 130:93-107(1993).
[11]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=2389555; PubMed=2389555;
RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.-O.O., Simon-Mateo C.,
RT Vinuela E.;
RL "Sequence and evolutionary relationships of African swine fever virus
RN thymidine kinase.";
RN virology 178:301-304(1990).
[12]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93281390; PubMed=8506138;
RA Yanez R.J., Bourisnell M., Nogal M.L., Yuste L., Vinuela E.;
RT "African swine fever virus encodes two genes which share significant
RL homology with the two largest subunits of DNA-dependent RNA
RN polymerases.";
RN Nucleic Acids Res. 21:2423-2427(1993).
[13]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93353606; PubMed=8102411;
RA Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
RT "African swine fever virus encodes a CD2 homolog responsible for the
RL adhesion of erythrocytes to infected cells.";
RN J. Virol. 67:5312-5320(1993).
[14]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94085774; PubMed=8262374;
RA Yanez R.J., Rodriguez J.M., Bourisnell M.E.G., Rodriguez J.F.,
RT Vinuela E.;
RL "Two putative african swine fever virus helicases similar to yeast
RN D6R.";
RN Gene 134:161-174(1993).
[15]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90223993; PubMed=2327074;
RA Lopez-Otin C., Freije J.M.P., Parra F., Mendez E., Vinuela E.;
RT "Mapping and sequence of the gene coding for protein p72, the major
RL capsid protein of African swine fever virus.";
RN virology 175:477-484(1990).
[16]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94123986; PubMed=8293992;
RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
RT "The DNA polymerase-encoding gene of African swine fever virus:
RL sequence and transcriptional mapping.";
RN Gene 136:103-110(1993).
[17]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93327788; PubMed=8335009;
RA Simon-Mateo C., Andres G., Vinuela E.;
RT "Polyprotein processing in African swine fever virus: a novel gene
RL expression strategy for a DNA virus.";
RN EMBO J. 12:2977-2987(1993).
[18]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93232310; PubMed=8474154;
RA Prados F.J., Vinuela E., Alcaml A.;
RT "Sequence and characterization of the major early phosphoprotein p32
RL of African swine fever virus.";
RN J. Virol. 67:2475-2485(1993).
[19]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=92260660; PubMed=1583732;
RA Alcaml A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.P.,
RN Carrascosa A.L., Vinuela E.;

```
RT "amino acid sequence and structural properties of protein p12, an
RT African swine fever virus attachment protein.";
RL J. Virol. 66:3860-3868(1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174976; PubMed=8438592;
RA Yanez R.J., Vinuela E.;
RT "African swine fever virus encodes a DNA ligase.";
RL Virology 193:531-536(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174941; PubMed=8382399;
RA Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
RT "African swine fever virus guanylyltransferase.";
RL Virology 193:319-328(1993).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=95159428; PubMed=7856088;
RA Simon-Mateo C., Freije J.M., Andres G., Lopez-Otin C., Vinuela E.;
RT "Mapping and sequence of the gene encoding protein p17, a major
RT African swine fever virus structural protein.";
RL Virology 206:1140-1144(1995).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=92263807; PubMed=1316688;
RA Garcia-Beato R., Freije J.M.P., Lopez-Otin C., Blasco R., Vinuela E.,
RA Salas M.L.;
RT "A gene homologous to topoisomerase II in african swine fever virus.";
RL Virology 188:938-947(1992).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=94091056; PubMed=8266720;
Query Match 3.8%; Score 114; DB 12; Length 1055;
Best local Similarity 19.2%; Pred. No. 1.1; 214; Indels 134; Gaps 25;
Matches 104; Mismatch 90; Mismatches 134; Gaps 25;
QY 15 TTGGSKSVNSELLO-MKYLSSGTMKLTFTTCLIVPSLVAF-----SMIFHOH 65
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
257 TTPSSCDTIQOSELEVKF--AGTLCKNHF--LRVHNILVFKRTSPSYCEICKRMH 311
QY 66 PPSNRMGAERAVLDAGVFPVNTINSKLLIGLIA--SGFDE--DSCLSR---YQ 116
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
312 DKNTLTLL-----RVGNKRYOHCRRDKNHSLMGSLSGTTFEVEYVVDQVMTKSEIVHE 366
QY 117 SVHYRKPSPYKPPSSYLISKLRNFE-----KLHKRGPGTESYKAKQLDQEHIDGG 169
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 SILFEELPDTQKHTIYDESSKRETERVPTLVYKAKMKIGKTVOALNTYLOKYGNNSISKQ 426
QY 170 ECKYVW--ISFSGLNRLTSLASVFLYALLT-----DRVLLVDRGKDDDFCEPFLG 221
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
427 TIRFVFRQIFSKNIQSR--LPNFTLYSEVTGDDLSYEKIVY-----QVESL-----FRL 474
QY 222 MSULLPLDPRMTQDFGL-NOESSRCYGM-----VKNOVIDTEGTLSHLYH 268
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
475 TSTAEPDLDLIDVESIFNOFNSGLHKYFAPSFALFMMLLETANVYICIDAMIGNRTYN 534
QY 269 LVHVDY-GD-----HDKMFEFGDQTFIGKVPMLVKINDNFVPSLMLIPG- 312
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
535 ILDFRGDVPITFHMNOYKRAQHDITYFTSSRETWLNLLKDLLEKKIVIPNSLMEAR 594
QY 313 -FDDELKLFEPQATVFH-----HLGRVLFHPTNOVGLV-----TRYEAVLSHADEK 360
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
595 ILQSFQKRPKEKKIGFYSSKTAHERESHFNVSYYWGLVDLLITPTISAGVSTEDKR 654
QY 361 IGIVRVFDEDDPQFQVMDQISSCQKELDPEVDTLVRSRHYVTPKHAKVALVYSLNA 420
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
655 -----FDVLVYGFNNMSCDVEFCQ-----MLGRVRELKSKCYKICLQCKQNV 697
```

```
QY 421 GVAENLK-----SMYWEYPTSGELIIGHOPSQEYQOTEKKMN 460
DB 698 -YPTETIDIMFTLQKRDILPOTINNHOLSFTYSKETGRPVYKTPPYHMLTMRIOHL 756
QY 461 GK 462
DB 757 SK 758
RESULT 5
Q92W93 PRELIMINARY; PRT; 2149 AA.
ID 092W93
AC 092W93;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F5A8.5 protein.
GN F5A8.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBETA;
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,
RA Yu G., Ecker J., Theologis A., Davis R.W.;
RA Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
SR EMBL: AC004146; AAD10656.1; --
SQ SEQUENCE 2149 AA; 237744 MW; BDE94CF7FC307B9B CRC64;
```

```
Query Match 3.7%; Score 112.5; DB 10; Length 2149;
Best local Similarity 23.6%; Pred. No. 4.2;
Matches 123; Conservative 65; Mismatches 186; Indels 147; Gaps 28;
QY 52 FSVLVAFSMITHOPBDSNRIMGFAERAVLDAGVFP-NVTNINSKUL-----L 98
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1351 FAVLVLFQ--GMHPASSMQVLPFSSAKIKSGDSSINSSFGKLTYELALPVFQSLSA 1407
QY 99 GGLLASGF-DEDSCLRYQSVNHRKSPYKPPSSYLISKLRNEKLHKRGPGTESYKRL 157
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1408 GRFFSSGFSLDLCQELDLYLST---SFHDSMDILAVSVQOL-----VGTKIYSTV 1459
QY 158 KOLDQEHIDGDECKYVWVISEFGLGNR-----ILSLASVFL-----YALTPRVLL 204
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1460 -----GEIISYPCFCYPAITGSPDIMEFWIYLPDFEFLRLVMPYALEKKRVL- 1507
QY 205 VDRGKMDLFCBPFLGMSMLPLDFPMTDQFDGLNOESSRCYGVWV-----NOVIDTE 259
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1508 -----MKFTF--FOIISONCPKDFLESEFAVSTIE--LCIGYLFKILHHRNEISPD 1556
QY 260 G-----TLSHLYL-----HLVHYGDHDKMFCBEGQTFIGKVPMLIVTNTYFVS 306
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1557 GIWDMNLSPFLFISIKTLVKRFELKRLNSAPLAFLLSGKC-IROVP-----TDAILPKA 1610
QY 307 LMLIPGFD--ELNKLFPQK-----ATVFHHLGYLFLHPTNOVGLVTRYEAVL 354
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1611 LEIVKSTNDLLELTPRASSQKPYTQGTNFAADSGHL-RAIFGACLIHMGDLTR----- 1663
QY 355 SHADEKIGIVRVFDEDDPQFQVMDQISSCQK-----EKLPEVDTLVRSRHYVTPKH 410
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1664 ---DCINGIQL-VDSKRSGLRKLQKLVECLEQVLSLAKLAYEFDCPYDER--NTNSI 1716
QY 411 KAVLYVS-----LNGYAEMLKSMWEYPTS-----TEGIIGHOP 446
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1717 CIYMLKSCQISIAAVVKDSNVQVATVLOVLSLVQRYNNPEKSFVILFVGLIGDIXP 1776
QY 447 SOEGYQOTEKKMNGKALAEYLL--SLTDNLVYSANSTF 484
DB 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

Db 1777 -----VNESVVIAGECLRFIMLQTHSITDELQKGFMSLF 1812
RESULT 6
ID 0997F0 PRELIMINARY: PRT: 2244 AA.
AC 0997F0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Polymerase (RNA polymerase).
GN L.
OS Nipah virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae.
OX NCBI_TextID=121791;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20287714; PubMed=10827955;
RA Chua K.B., Bellini W.J., Rota P.A., Harcourt B.H., Tamlin A., Lam S.K.,
RA Ksiazek T.G., Rollin P.E., Zaki S.R., Shieh W., Goldsmith C.S.,
RA Gubler D.J., Roehrig J.T., Eaton B., Gould A.R., Olson J., Field H.,
RA Daniels P., Ling A.E., Peters C.J., Anderson L.J., Mahy B.W.;
RT "Nipah virus: A recently emergent deadly paramyxovirus.";
RL Science 288:1432-1435(2000).
[2]
RP SEQUENCE FROM N.A.
RA Harcourt B.H., Tamlin A., Halpin K., Rollin P.E., Ksiazek T.G.,
RA Bellini W.J., Rota P.A.;
RT "Molecular characterization of the Polymerase Gene and Genomic Terminal
RT of Nipah Virus.";
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=UMMC2;
RX MEDLINE=21405817; PubMed=11514724;
RA Chan Y.P., Chua K.B., Koh C.L., Lim M.E., Lam S.K.;
RT "Complete nucleotide sequences of Nipah virus isolates from
RT Malaysia.";
RL J. Gen. Virol. 82:2151-2155(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=UMMC2;
RA Chua K.B., Chan Y.P., Koh C.L., Lim M.E., Lam S.K.;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF212302; AAK29089.1; -
DR EMBL: AY029768; AAK50555.1; -
DR InterPro: IPR001016; Viral_RNA_pol_L.
DR Pfam: PF00946; Paramyx_RNA_pol. 1.
SQ SEQUENCE 2244 AA; 257230 MW; 1DE123E651A738C9 CRC64;

Query Match 3.7%; Score 112.5; DB 12; Length 2244;
Best Local Similarity 18.8%; Pred. No. 4.5;
Matches 104; Conservative 79; Mismatches 192; Indels 177; Gaps 25;

OY 1 MDONSRRSSPRLRTTGGSKSVNFSELLQMKYLSGTMKL--TRFTTCLIVFSVLAF 58
Db 1804 VENHKYRRIG---LNSSCYKALNLSPLIQ-RYLPFGAQRRLFGESGSMMLYOSTLQ 1859
OY 59 SMIFHQHPSDNSNRIMGFARVLDAGVFNVTNINSK-----LLGGLASGDESDCSLS 113
Db 1860 SISFYNSGIDGYIPQORELK-----LFPSEVSIADSDPSLTGKLGVLVPLFNG----- 1909
OY 114 RQSVHYRRKPSPYKSSYLISKRYNEKLHKRCGPETSEYKALKKLDDEHIDGDECKY 173
Db 1910 -----RPETTMIGNDSYEYIINR---TAGRSIGLVSDME----- 1942
OY 174 VVWISFSGLGNR-----LSLASVFLYALTDRLVLDVDRGMDMDLCEPFLGMSWL 226
Db 1943 -----SGIDKNVEELIVESHLSIAINVMMDGL-----VSKIATYTPGFPI SRL- 1988
OY 227 PLDFPMTDFDGL-----NQESSRCYGYVKNQV---IDTECTLHLVYLH-LVHDY 273

```

```

Db 1989 ---FNNRSYFGLVLCFPPVYSNPDSLEYVLLCQTKVTIYPPQKVLHSHNLHDEVNDQ 2045
OY 274 GDHDKMFEGEGDTFFIGKVPWLIVKTDNKFVPSYMLIPGEDDLANKLFQKAVFNHLGR 333
Db 2046 GITSVLEFKINSQS-----KQFHDLKR-----VYQIDQ 2074
OY 334 YLFHPT-----NQW---GLVTRYEAYLSHADEKIGIVRV-----FDE 370
Db 2075 PFEVFPKITSDEYVLLQAGLKLNGPRELISEISYDGSIDINTLRDTIITMLNEAMVFPD 2134
OY 371 DRGPFQHV-----MDQISSCTQKE---KLLPEVDTLVRSRHVTPRKAVLVY 416
Db 2135 NRSPSHLEPPYVLETRRIKTIWNCVTKKVIYVSLIKPDKTSSELYHKNNIRKVLTI 2194
OY 417 SLNAGLAENLKSMTWEPYPTSGEIIIGVHQPDSQCYOOTEKKMHNGALAEMLTSLTDNI 476
Db 2195 DFRS-----KIMTKTLP-----KGMQERREK--NG--FKEWIVLDSNRE 2230
OY 477 VTSAMSTFGYVA 488
Db 2231 VKIWMKIIGIYS 2242

RESULT 7
ID 091AE5 PRELIMINARY: PRT: 2244 AA.
AC 091AE5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE RNA polymerase.
GN L.
OS Nipah virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae.
OX NCBI_TextID=121791;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=UMMC1;
RX MEDLINE=21405817; PubMed=11514724;
RA Chan Y.P., Chua K.B., Koh C.L., Lim M.E., Lam S.K.;
RT "Complete nucleotide sequences of Nipah virus isolates from
RT Malaysia.";
RL J. Gen. Virol. 82:2151-2155(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=UMMC1;
RA Chua K.B., Chan Y.P., Koh C.L., Lim M.E., Lam S.K.;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY029767; AAK50546.1; -
DR InterPro: IPR001016; Viral_RNA_pol_L.
DR Pfam: PF00946; Paramyx_RNA_pol. 1.
SQ SEQUENCE 2244 AA; 257290 MW; 993DC4782846E268 CRC64;

Query Match 3.7%; Score 112.5; DB 12; Length 2244;
Best Local Similarity 18.8%; Pred. No. 4.5;
Matches 104; Conservative 79; Mismatches 192; Indels 177; Gaps 25;

OY 1 MDONSRRSSPRLRTTGGSKSVNFSELLQMKYLSGTMKL--TRFTTCLIVFSVLAF 58
Db 1804 VENHKYRRIG---LNSSCYKALNLSPLIQ-RYLPFGAQRRLFGESGSMMLYOSTLQ 1859
OY 59 SMIFHQHPSDNSNRIMGFARVLDAGVFNVTNINSK-----LLGGLASGDESDCSLS 113
Db 1860 SISFYNSGIDGYIPQORELK-----LFPSEVSIADSDPSLTGKLGVLVPLFNG----- 1909
OY 114 RQSVHYRRKPSPYKSSYLISKRYNEKLHKRCGPETSEYKALKKLDDEHIDGDECKY 173
Db 1910 -----RPETTMIGNDSYEYIINR---TAGRSIGLVSDME----- 1942
OY 174 VVWISFSGLGNR-----LSLASVFLYALTDRLVLDVDRGMDMDLCEPFLGMSWL 226
Db 1943 -----SGIDKNVEELIVESHLSIAINVMMDGL-----VSKIATYTPGFPI SRL- 1988

```

```

0Y 227 PLDFPMTQDFGL-----NQESSRCGYAWKQV---IDTBGTSJLYLH- LVHDY 273
      : : : : :
Db 1989 ---FMWYSXGYLEVJCFPVVYNSNDSTEVYLLCLOKTIVTPIVPOKVLHSHSLHDEVDNQ 2045S
      : : : : :
0Y 274 GDHDKMFCEDDOJFICKVPMVLKVDNTPVPSMLNIPGFDDELKMLPFQOKATVFNHNGR 333
      : : : : :
Db 2046 GITSYTFKIKNSQ-----KQFHDLKK-----YQIDQ 2074A
      : : : : :
0Y 334 YLFHPT-----NQWV--GLVTRYEAYLSHADEKIGIOVR-----FDE 370
      : : : : :
Db 2075 PEFVPTKITSDEQVLLDAGLKLNGPFLIKSEISYDIGSINTLRDTIIMLMEANNYVDD 2134A
      : : : : :
0Y 371 DPGPFOV-----MDQISSCTQK---KLPEVDLVERSRYHNPCKHKAVALVT 416
      : : : : :
Db 2135 NRSFHNLEPYPVLEFRTIRIKTIMNCVTKKVIYSLIKFDDTSSSELYHNKNNIRKXVLL 2194A
      : : : : :
0Y 417 SLNCGYAEMLKSMWEPTSTGELIGVHDPDSOEGYQOTFKKHNKGALAEWYLSLTDNL 476
      : : : : :
Db 2195 DFRS-----KLMKTLR-----KGMEREREK--NG--FREYVIVDLSRE 2230A
      : : : : :
0Y 477 VTSASMTGUYVA 488
      : : : : :
Db 2231 VKIWMKIIGIYS 2242

```

RESULT	8
085712	
ID	085712
AC	085712;
DT	01-NOV-1998 (TREMBLrel. 08, Created)
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	Modulation protein NodZ.

OS Rhizobium fredii (Sinorhizobium fredii).
OG Plasmid pSFHH103d.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_taxid=380;
[1]
SEQUENCE FROM N.A.
RN STRAIN-HH103.
RP MEDLINE=99164952; PubMed=10065558;
RC MEDLINE=99164952; PubMed=10065558;
RX MEDLINE=99164952; PubMed=10065558;
RA Lemarabet Y., Bellogin R.A., Cubo T., Espuny R., Gil A., Krishnan H.B.,
RA Megias M., Ollero F.J., Pueppke S.G., Ruiz-Sainz J.E., Spaink H.P.,
RA Tejero-Mateo P., Thomas-Yates J., Vinardell J.M.;
RT "Mutation in gbp-fucose synthetase genes of Sinorhizobium fredii alters
RT nod factors and significantly decreases competitiveness to nodulate
RT soybeans.";
RT Mol. Plant Microbe Interact. 12:207-217(1999).
RL FMBL: AF072888; AAC27749.1; -.
KW Plasmid.
KW SEQUENCE. 322 AA; 36535 MW; 0023D958F17781CB CRC64;
SQ

Query Match	3.7%	Score 110.5	DB 2	Length 322
Best Local Similarity	20.2%	Pred. No. 0.37		
Matches 74; Conservative	51; Mismatches	109;	Indels 133;	Gaps 17

```

QY 172 KYVWVISPSGLGNRIILSLASVEFYALLTDRVLVD-RGKMDLDFCEPFLGMSWLPDF 230
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 4 RYVLSRRRTGCGDGLWLSLAAMRYAQATARIANDWMS-----CY-----LDQ 47
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 231 PMTDQFGLNDESSRCYGVYMKNOVIDTEGTLSHLYLHLVHDYGDHDKMFECEGDTFIG 290
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 PFTNAFP-----VFPEPKD-----IA 64
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 291 KYVPLIYKTDNY-----FVPSLLILGFD-----DELKFLPQKATV- 327
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 65 GVPEI---CDNQINEFSFGPEFFPMNNKPAIECVYPPRAQVFRERDELDELPOAQDVE 121
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 328 -----FHHLGRLF--HPTNQWGLVTRYEATLSHADEKIGIQVRYEVE 370
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Db 122 ANTVVCADCLMKRCDEBEEROLFCVSKRRAETQARIDALIOEHF-YGYSIAIGVHVR---- 176
Qy 371 DGPQPHVMDIOISSCTQREKLLPEVDLTVERSRHVNPCKHKAVALVTSINAGYAEULEKSMY 430
   : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 -HGNEDEDVMDIAPYADPDLVAHQYCTAINAKKALPHKPKPYRVLICLDSARVLDQVSSRF 235
Qy 431 WEYLP-----STGELIIVHOPSDQEGYQOTEKKMHNG-KALAEWYLLSLTGNLV---- 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 PDLLEIIPKSFRAQDQGPL--HSP-----DLVGEGISALVEMVLLIGCTPIVIFPP 284
Qy 478 TSANSTF 484
   ||| : :
Db 285 TSAFTFY 291

```

```

RESULT 9
ID 021310
AC 021310 PRELIMINARY; PRT; 529 AA.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE k08C7.5 protein.
GN k08C7.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Betks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RL EMBL: Z70286; CAAG4290.1; -.
DR InterPro: IPR000960; FlyBase:conf_moxnog.
DR Pfam: PF00743; FMO-1like; 1.
SQ SEQUENCE 529 AA; 60140 MW; 05EDD3446D3FAAC1 CRC64;

```

Query Match	3.78;	Score 110;	DB 5;	Length 529;
Best Local Similarity	20.38;	Pred. No. 0.85;		
Matches 102; Conservative	78;	Mismatches 142;	Indels 180;	Gaps 30;

[illegible]

```
Db 290 RIANGTVRIKPIKIKKFDYAIHFEEDGTVPHVDEYVMSTGSEFEENLIEHGKLVSENE 349
Oy 399 VERSHH---VMTPKHKAIVLTSLNAGY-----AENLSKMVEYFTSGEIGGVQPSQE 449
Db 350 VDLFEYFMPVATSDHNSLCITIGLIPGSEIMPVSFOARVEFANNVSGNNL----PKRS 405
Oy 450 GYOOTEKRMHNGKALAEWYLLS 471
Db 406 --QMSEDVLNKKREMAAQOFVKS 425

RESULT 10
ID 025252 PRELIMINARY; PRT: 658 AA.
AC 025252:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein HP0513.
GN HP0513.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khaliq M.G., Glodek A.,
RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL; AB000565; AAD07579.1; -
DR TIGR; HP0513; -
DR InterPro; IPR004919; DUF262.
DR Pfam; PF03235; DUF262; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 658 AA; 78181 MW; 72CF7028B68912D9 CRC64;

Query Match 3.7%; Score 110; DB 16; Length 658;
Best Local Similarity 17.8%; Pred. No. 1.2;
Matches 102; Conservative 79; Mismatches 191; Indels 202; Gaps 23;

Oy 37 GTMLTRFTTCLIVSVLAFSMIFHOHPSDSNRIMGFARVLADGVFPNVTININDK 96
Db 28 GOQRJT-TITILLIA-----LRNHLSSEVEILEKFSRKELESYLIN--SKDDDK 74
Oy 97 LLGSLASGFEDSCLSRYQSVHRYKRPSPYKSSYLLSKLNRYELHRCRGPSTGYKKA 156
Db 75 KF-RLISESKQDLISLIDK-NKRKPS-EPYKIVNFELPEKMWISENDKLETFIKG 130
Oy 157 LKOLDQEHIDGDECKYVWIS-----FSGLGNRILSLASVFLVALLDTRVLIV 205
Db 131 LKLL-----MIVWISLIDKGDPPQLIFESMNSKDIELTQT--DLIRNYIME 175
Oy 206 DRGDMODLPEPFLGMSWMLPLDFPMTDQF----- 236
Db 176 TEVEQDEYFYQVYRAMEREFOETLEFRVRYHLTIKIKIPNEKRYEAFDYROKK 235
Oy 237 ---DGLNOESSRCYGVW-----KNQVIDTEGTLISH-----LYLHIVHDYGDH 276
Db 236 GIEIEDLLKDLQKICGICQIAFKKEDDKLANKALSFLVNLMDVITYPLLELSDYKD- 294
Oy 277 DKMFCEGDQTFIGKVPWLIVKTDNYFVPSLMLPGF-----DDELINKLFP----- 322
```

```
Db 295 -----GLSKQD--FPIPIYLIISYICRAVCGLGNSLNKVPSPSTK 335
Oy 323 --QKATVPHHLCRYLF-----HPTNOYGLV-----TRYEAYLSHADEKIG 362
Db 336 HIOKDEYFKSLKAHFVCLTEKORFPNNDFFKFLFTIDYFKFKKKKFLERLENEDJKEP 395
Oy 363 IQVRFEDEDPGFQHVMOIISCTQEKILPEVDLVERSRIV--NTPKHKAIVLTSLNA 420
Db 396 V-----DTQKNIEMHMQTLTPWQRDGENFQAHEKYLHTIGN-----LTLLGYNS 444
Oy 421 GYAEN-----LKSM-----YMEYPTST 437
Db 445 KYSNNSFOEKRMEKGFQSSLSKLNQSLKDLSPGEKELEKASDLADMAKIMYTPILLE 504
Oy 438 GEIGVHQPDSQEGYQOTEKRMHNGKALAEWYLLS 471
Db 505 AETLEEYKPKKKEKKEKKEEYKLEKKKVVDLIS 538

RESULT 11
ID 007183 PRELIMINARY; PRT: 1102 AA.
AC 007183:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE Putative origin of replication binding protein lwm6dl.
OS African swine fever virus (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10497;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=93275762; PubMed=8389041;
RA Sussman M.D., Lu Z., Kutish G.F., Alfonso C.A., Rock D.L.;
RA "The identification of an African swine fever gene with conserved
RT helicase motifs and a striking homology to herpes virus origin binding
RT protein, UL9."
RL Nucleic Acids Res. 21:2254-2254(1993).
CC - FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
DR EMBL; L12174; AAA42719.1; -
KW DNA replication; DNA-binding; Helicase.
SQ SEQUENCE 1102 AA; 129255 MW; 4B06FA67304C60 CRC64;

Query Match 3.7%; Score 110; DB 12; Length 1102;
Best Local Similarity 19.0%; Pred. No. 2.6;
Matches 103; Conservative 94; Mismatches 210; Indels 136; Gaps 26;

Oy 15 TTTGSKSVNSELQ--MKYLSGTMKLTFTTCLIVSVLVAF-----SMIFHOH 65
Db 251 TTSSSCDSIQOSELEEVLF--AGTLCKNHCF--LRVYKNLVLEKRRSPSYCEICRBMH 305
Oy 66 PSDSNRIMGFAEARVLADGVFPNVTININDKLLGLSLA--SGFDE--DSCLSR----YQ 116
Db 306 DKDNTFLI-----RTGKVKYQHCHDKHSHLSMGSIGTNNFVTTYEQVMSKIEVNE 360
Oy 117 SVHYRKPSPYKPSYLLSKLNRY-----KLHKRCGPSTGYKKA LKOLDQEHIDG 169
Db 361 SILFEELPDQTKHIYDESSMREYERVPLLVKAKKIKGTIQ-----LRNYLQKYGDS 415
Oy 170 ECKY--VWVIFSFGIGNILS--LASVFLXALLT-----DRVLLVDGKMDMDLFCPEP 219
Db 416 ISKQOTIRFVTRQIFSNIOQTRLPNFTLYSEVTDGDDSYERVII--OVESL---F 466
Oy 220 LGMWLPPLDFPMTDQFQGL--NOESSRCYGVW-----YKNQVIDTEGTLISHLY 266
Db 467 RLTSAPRYVDLLIDVESIRFNGFSGLHKYFAPSFALPMMLERANVITCLDANLKRT 526
Oy 267 LHLVHDY-GD-----HDKMFCEGDQTFIGKVPWLIVKTDNYFVPSLMLIP 311
Db 527 YNIQRFRGDVPIFFHNMVQYKQANDMYFTSSREIMWNLMLKDLLEDKKLVIPNLSLME 586
```

QY	312	C	FPDELINKLEPRKATVH-----HLCRYLFHPINOWGLV-----TRYEATYSHAD	358
QY	587	ARLLQTFIOKKPEKKIGFYSKSSAHERESHFNNSYWGILDLITYPTISAGSYED	646	
QY	359	EKIGIVRVEDDDPGFQVHMDISCTQ-----KEKLLPEVDLTV	399	
Db	647	KR-----FDVLVGFNNMAGCDVEHCCOMGLRGVRLKSKCYKICLQGGONPPET---	696	
QY	400	ERSRHVTPKRAVLVTSINAGYAENLKSMEWEYPTSTGEIIGVHOPSGEYQOETKKMH	459	
Db	697	EDIEHFTLQKRPDLTFLQTSIN-----HQLSTFYCKETGRPIYKPKPYHLMLETWRIQH	749	
QY	460	NGK 462		
Db	750	ISK 752		
RESULT 12				
ID	09KRO1	PRELIMINARY;	PRT; 567 AA.	
AC	09KRO1			
DT	01-OCT-2000	(TREMBLrel. 15, Created)		
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)		
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
DE	Catalase (EC 1.11.1.6).			
GN	VC1585.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_taxid=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EL TOR N16961 / SEROTYPE O1;			
RX	Medline=20406833; Pubmed=10952301;			
RA	Hedlberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,			
RA	Dodon R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,			
RA	Ernolaeva M.D., Yamethayan J., Bass S., Qin H., Dragol I., Sellers P.,			
RA	McDonald L., Utterback T., Felschmann R.D., Nierman W.C., White O.,			
RA	Saizberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RT	cholerae";			
RL	Nature 406:477-483(2000).			
CC	-!- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND			
CC	SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN			
CC	PEROXIDE (BY SIMILARITY).			
CC	-!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.			
CC	EMBL: AF004235; AAF94739.1; -.			
DR	HSSP: P00432; 4BLC.			
DR	TIGR: VC1585; -.			
DR	InterPro: IPR002226; Catalase.			
DR	Pfam: PF00199; catalase; 1.			
DR	PRINTS: PR00067; CATALASE.			
DR	ProDom: PD000510; catalase; 1.			
DR	PROSITE: PS00437; CATALASE_1; 1.			
DR	PROSITE: PS00438; CATALASE_2; 1.			
KW	Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase;			
KW	Complete proteome.			
SO	SEQUENCE 567 AA; 6409 MW; 5EDA31B2501A4A21 CRC64;			
Query Match 3.6%; Score 107; DB 16; Length 567;				
Best local Similarity 22.1%; Pred. No. 1.7;				
Matches 100; Conservative 56; Mismatches 144; Indels 152; Gaps 26;				
QY	32	KLSSG---TKMTGTFPTCLIVFSVLAFSMIFHQ--HPSDSNRKIMGFAEARVADAGVF	86	
Db	145	EVVAGDPSDLTSLAPFTSKGITPEVFAFRFTVHISKSGPEFLRDPGRATKFTYEG--	202	
QY	87	PNVTVINSIDKLGLGILLAGGFDEDSCLSRQSVHYKRPSP-----YKPS-	130	
Db	203	-----NMD-LVGNMLPVFLTRDS-IRPRDWHSLAKPSVYTMQDPNRFDFFSHPGSGT	254	
QY	131	-----YLISKLRNYEKLRKRCGPFTESYKALKOLDQEHIDGDECKYVV--WISFSGIGN	184	

[illegible]

RESULT 14
094648 PRELIMINARY: PRT: 3844 AA.
AC 094648;
DT 01-FEB-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE AARPL protein (Fragment).
CN AARPL.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=97378065; PubMed=92347746;
RA Barale J.C., Candelle D., Altal-Bonnefoy G., Dehoux P., Bonnefoy S.,
RA Barale J.C., Pereira da Silva L., Langsley G.;
RT "Plasmodium falciparum AARPL, a giant protein containing repeated
RT motifs rich in asparagine and aspartate residues, is associated with
RT the infected erythrocyte membrane."
RL Infect. Immun. 65:3003-3010(1997).
DR EMBL: Y08926; CAAT0130.1; -
DR InterPro: IPR001841; znf_r1ng.
DR SMART: SM00184; RING; 1.
FT NON_TER 1
FT SEQUENCE 3844 AA; 456075 MW; 23D0233261C30DBB CRC64;
SQ

Query Match 3.5%; Score 106; DB 5; Length 3844;
Best Local Similarity 18.1%; Pred. No. 37;
Matches 91; Conservative 60; Mismatches 131; Indels 220; Gaps 24;

QY 23 VNESELLQMKYLSG-TWKLTRTTTCLIVFSVLVAPSMIFHQPDSNRIMGFAEARVL 81
DB 371 INOKRIILOERENGNGNMKIT-----LFHFKEKF-----TSN-----DDIYL 407
QY 82 DAGVPPNTNNTNSKLLGLGSLASGFDESDCLSRQSVHYRKPSPYK-----PSSYLSK 135
DB 408 ESNFYKMTMNNSSSKL-----RECNNKYR-IKYRETIQEKNIKIKCKNIYIEKK 456
QY 136 LRN--YEKLHRCGPGTESYKALKQDOEHIDGDECKYVWISFSLGNRIILASVF 193
DB 457 LKNSKE-----FNEPLD-----IKY----- 472
QY 194 LYALLTRVLLVDRGKMDLFCPEFLGMSWLLPLDFPMTDQFDGLNDESSRC----- 246
DB 473 -----DKEKNITLHP-----LYIDLVMGDIFFELNLFDLKLNILFI 511
QY 247 -----YGYMKKNQYIIDEBSLHLYLVHVDYGDHDKMFCFEGQOTPIGKVPW 294
DB 512 NIQLTFEYIEIKYQWNGP-----OMIFQVGEYDSLFQND--LIG-IQF 556
QY 295 LIVTDFVYFVSLMLPGFDEDLNKLFPQKATVFHHLGRYLFHPTNOVWGLVTRYEA-- 352
DB 557 ALIMMN-----VLPLEYENKRYLLDSHVNLPH-----QIMNIKVOYDEPLK 597
QY 353 -----YLSHADEKIGIOVRVEDDEPGRFOHVMDOISSTQCEKLLPEVDY 397
DB 598 KLPMKSNKYNDVQDYIHKRDEICINKEYDEDTNKKY-----IDT 638
QY 398 LYEKSRHYNTPKHKAVLVTSNAGYAENLKSMMY--EYPTSGEIIIVHQPDSOG--Y 451
DB 639 SPNFSYNHNIQNHVY-----EEHSEDKRPYNNKIKYIKKNDFFEEHKKMYESMLLY 692
QY 452 QOTEK-----KMHNGK 462
DB 693 NRSKKSNDNNTLPLVNNKMGNSK 714

RESULT 15

001660
ID 001660 PRELIMINARY: PRT: 381 AA.
AC 001660;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 44.4 kDa protein.
CN T28F2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Madsen C., Fronick B.;
RT "The sequence of C. elegans cosmid T28F2."
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF000198; AAB53053.1; -
DR InterPro: IPR002516; GT_11.
DR Pfam: Pf01531; Glyco_transf_11; 1.
DR Hypothetical protein.
SQ SEQUENCE 381 AA; 44370 MW; 502A9742034053C1 CRC64;

Query Match 3.5%; Score 105; DB 5; Length 381;
Best Local Similarity 18.3%; Pred. No. 1.4;
Matches 67; Conservative 49; Mismatches 97; Indels 154; Gaps 18;

QY 296 IVKTDNYFVPSL-----WLIPGF-----DDELNKLFPQKATVF 328
DB 11 ILKNNKEFIYIIVLLVIFOMWVPGYDYDNOISTLYYCDKNPSDQRIYLFPRVTVTF 70
QY 329 H-HLGRYLFHPTNOVGLV-----TRY--EAYLSH-----ADEKIGIO 364
DB 71 DGLGNQLF-----EVFSLGLALKLNRPALFNSEDMILSKNLLREQVPOVAERVISIP 126
QY 365 VRV-----FDDEPGRFOHVMDOISSTQKE-----KLLPEVDLYEKSRHYNT 407
DB 127 IEIPSSRFYSPACHPQFSPSLSCESQKYLVIDGHYFQSHKVFSSISITSIR--KWLNP 184
QY 408 PKHKA-----VLVTSNAGYAEN-----LKSMMYEP-- 434
DB 185 PQEKKLLKMKIRKDERFRKICVHIRGDFLTDSQAHGTSNFTIRAVDHLTYOHPGLV 244
QY 435 ---TSTG---EIIGVHQPDSOGYQOTEKMHNGKALAEMLLSTLDN--LVTSAMSTFG 485
DB 245 YLEFNDPQWREKTAANL-----DYQSDVKVMTSEAIKDLVFAQIHCDAVLITAPSTFG 300
QY 486 YVAQGLGLKMWILYRPNRTTPRPSGCRASMECFHSPFPYDCKATGTDIGLVPHV 545
DB 301 W-----WLGYSKNSQSSQTPNGR-----GFLPVHV 326
QY 546 RHCEDIS 552
DB 327 EIGDVT 333

Search completed: February 27, 2003, 15:31:50
Job time : 46 secs

```
Db 243 SLIDICGATILMTADGLFAPFQKHTEALRQEIINARPCGCVGLNTLAFPSINRREVE 302
OY 431 ----WEPTSTGELICGH 444
Db 303 EKOPWAY-ISCGHVGHY 319

RESULT 2
US-09-815-242-13660
; Sequence 13660, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13660
; LENGTH: 502
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13660

Query Match 3.0%; Score 90.5; DB 10; Length 502;
Best Local Similarity 18.7%; Pred. No. 2.9;
Matches 97; Conservative 94; Mismatches 176; Indels 151; Gaps 29;

OY 58 FSNIFHO-----HPSDS--NRIMGFAEARVLDAVFPN-----VTNINSDDLGLGLAS 104
Db 36 FLOIFPQAGWVHNANETWNSVSIACAFIESGVKPNQIEIGITNORETVV----- 89
OY 105 GDEDECSLSRYOSVHYRKSPYKPPSSYLLSKLRN---YEKLHRCGPGETESYKALK-QL 160
Db 90 -MDKRTGLPIYNAIYWOS-----ROTAPLAEQLKNOGYEKEFKEKGLIIDAFASTKYRW 144
OY 161 DQEHIDG-----DDEC-----KYVWLSFSGLGN--RILSLASVFIYAL-----LTDRV 202
Db 145 ILDHVEGAQERAKEKELFGTIDTWLWKLTDGAHVTDYSNAARTMLYNIKELKWDDEI 204
OY 203 L-----LVDRGKMDMLF--CEPFLGMSWLLPLDFPMTQ-----PDGLNQESSRCYG 248
Db 205 LFLINIPKAILLEVANSNSITIKTAPFHHYGGFVPISGMAGDQALALGQLAFEPG----- 260
OY 249 IWKKNQVITDEFTSLHLHLVHD-----YGDHDKMFFECGSDQTFI--GKVP 293
Db 261 -WKN-----TYGTSGFIIMNTGEMQISENNLLTITIGYINKKYVYALGSGIFIGSAIQ 315
OY 294 WL-----IVKT-----DNVFPSPSLMLLGFDDELNKLKLPQAKATVFNHILGRYLFPT 339
```

```
Db 316 WLROGLRWENPESREKRYARDSHNNDVYVPAFTGLGAPYNNQNA-----R 362
OY 340 NOVWGLVTR--YEAVALSHADEKIGIQRVFEDEGPPFOHVMDOISSCQKEKLEPEVD- 396
Db 363 GSVFGL-TRGTSKEDPIKATIGSIAYQVR-----DIIDTQOVDPQTAIQYLKADG 411
OY 397 -----TLVSRSHVNTPKKAVLVTSINAGIENLSMWEYPTSTGE 439
Db 412 GAAMNFMQFOADILIGIDIAKAKMLETTALGAFLAGLSVGWMDLDEL--KLINTEGE 469
OY 440 IIGVHOPSOEGYQOTEKKMHNG--KALAEWYLSLTDN 475
Db 470 LF---EPSMN--ESKRQQLYKWKKAIVKATQVFAEYDD 502

RESULT 3
US-09-898-533-3
; Sequence 3, Application US/09898533
; Patent No. US2002010656A1
; GENERAL INFORMATION:
; APPLICANT: Gemmill, Robert M.
; APPLICANT: Drabkin, Harry A.
; TITLE OF INVENTION: TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR,
; TITLE OF INVENTION: PATCHED
; FILE REFERENCE: 93445-00004
; CURRENT APPLICATION NUMBER: US/09/898,533
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: US/09/268,140
; PRIOR FILING DATE: 2000-03-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 1286
; TYPE: PRN
; ORGANISM: Drosophila melanogaster
US-09-898-533-3

Query Match 3.0%; Score 89.5; DB 10; Length 1286;
Best Local Similarity 20.6%; Pred. No. 16;
Matches 64; Conservative 41; Mismatches 81; Indels 125; Gaps 17;

OY 218 PFLGMSW-----LPLDEPMTDOPDGLN-----QESSRCYGM 250
Db 672 PFLMSWVKFLVWVGFLAALLSSLYASTRLDGLDITDLYPKDSNEKFLDAQTRLEGFY 731
OY 251 VKNQYITDEGTLSH-LYLHLVHDYGDHDKMFFECGSDQTFIGKVPWLIVKTONYFVPSLML 309
Db 732 SMYAV--TQGNFEYPTQOOLRLDY--HD-----SFV-RVPH-VIKNDNGGLPDPFL 776
OY 310 IPGFDELNKLFPQKATVFNHILGRYLFPTQVWGLVTRYRYEALVSHADEKIGIQVYFD 369
Db 777 L-----LF-----SEWLGNIQKIFD 791
OY 370 EDPGPPQHVMDQISSCQKEKLEPEVDT-----LVSRSHVNTPKKAVLVTS--LNA 420
Db 792 E-----EYRGRRLTKECMPNASSDALIAYKLLIVQIGHVDPNDPKDELVLNRLVNS 842
OY 421 GYAEULKSMY-----WEYPTSTGELITGVHOPSQESYQOTEKKMHNGAL-----A 465
Db 843 DGIINRAFYVYVLSAMATNDVFAVGASQGL--YRPROYFHQPNFEYDLKIPKSLPLVYA 900
OY 466 EM--YIISLTD 474
Db 901 QMPYLIHGLTD 911

RESULT 4
US-08-954-701A-6
; Sequence 6, Application US/08954701A
; Publication No. US20030032085A1
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
```



```
Query Match 2.9%; Score 87; DB 9; Length 418;
Best Local Similarity 18.6%; Pred. No. 4.7;
Matches 90; Conservative 48; Mismatches 159; Indels 188; Gaps 22;

QY 99 GGLASGF-----DDSLSRQSHYKPKSPYKPSLISKLNRYEKLRRCGGTE 151
DB 18 GELIVLGYNSLPLNGDGRKRSRFALEKPRKANGVCPST-----VHIACTP--- 63
QY 152 SYKKALKODQEHIDGDECKYVWVISFGLGNRIILSLASVFLYALLTDRLVLYDGRKM 211
DB 64 QAKAISNMKQHSIS-----YTLSSRAQTVVVEYTHDS 95
QY 212 D-DLCEPFLGMSWLLPLDFPMTDOFDGLNQESSRCYGVWKNQVIDT---EGTLSHLYL 267
DB 96 NTDMF---QIGRSTESPIDEVVTDVTPGQSNS-----DTQSVOSTISRFAC 139
QY 268 HLVHDGDHDKMFECEGDOTFICKVPMILYKTDNYFVPSIMLIPGDDLEKLPQKATV 327
DB 140 RII-----CERNPFTART-----YVAFDSSKNFLIGERKAK 172
QY 328 FHHLGRLYFHPNTQVGLVTRYEAYLSHADEKIGIOV---RVFDED--PGPFOHV--- 378
DB 173 WK-----TSDQMDGLTTN-----GVLVHPRNGFTEDSKPGIMREISVC 212
QY 379 -----MDQISSCTQKEKLLPEVDY-----LYERSRHVNTPKHRAVL 414
DB 213 GNVSFLRETRSQAQGRKMV-EIETNOLQDGLIDLCGATLLMRTAEGLSHTPYVKHLEAL 271
QY 415 VTSLNAGYAENLKSMEWEPSTGEIIGVHPQSGYQOTEKKMHNGKALAEYLLSLTD 474
DB 272 RQEINARQ-----CPVGF-----NTLAFPSMKRKRDVVD 301
QY 475 NLVTSAMSTFGYVAGLGLKLPWILYRPENRTTPDPSCGRAMSMEPCFHSPEFYDCAKT 534
DB 302 EKQPMVYLLNCGHV---HGYNHW--GNKEERDQKDRCPMCRSVGP--YVPLMLGCEAGF 353
QY 535 GIDTG 539
DB 354 YVDAG 358

RESULT 7
US-10-041-030-2
; Sequence 2, Application US/10041030
; Patent No. US20020150934A1
; GENERAL INFORMATION:
; APPLICANT: Powers, Scott
; APPLICANT: Mu, David
; APPLICANT: Xiang, Phil
; APPLICANT: Peng, Yue
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Diagnosis and Treatment of Cancer Using Mammalian
; FILE OF INVENTION: Pellino Polypeptides and Polynucleotides
; FILE REFERENCE: 018781-006810US
; CURRENT APPLICATION NUMBER: US/10/041,030
; CURRENT FILING DATE: 2001-12-28
; PRIOR APPLICATION NUMBER: US 60/259,502
; PRIOR FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 2
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human pellino 1
US-10-041-030-2

Query Match 2.9%; Score 87; DB 12; Length 418;
Best Local Similarity 18.6%; Pred. No. 4.7;
Matches 90; Conservative 48; Mismatches 159; Indels 188; Gaps 22;
QY 99 GGLASGF-----DDSLSRQSHYKPKSPYKPSLISKLNRYEKLRRCGGTE 151
```

```
DB 18 GELIVLGYNSLPLNGDGRKRSRFALEKPRKANGVCPST-----VHIACTP--- 63
QY 152 SYKKALKODQEHIDGDECKYVWVISFGLGNRIILSLASVFLYALLTDRLVLYDGRKM 211
DB 64 QAKAISNMKQHSIS-----YTLSSRAQTVVVEYTHDS 95
QY 212 D-DLCEPFLGMSWLLPLDFPMTDOFDGLNQESSRCYGVWKNQVIDT---EGTLSHLYL 267
DB 96 NTDMF---QIGRSTESPIDEVVTDVTPGQSNS-----DTQSVOSTISRFAC 139
QY 268 HLVHDGDHDKMFECEGDOTFICKVPMILYKTDNYFVPSIMLIPGDDLEKLPQKATV 327
DB 140 RII-----CERNPFTART-----YVAFDSSKNFLIGERKAK 172
QY 328 FHHLGRLYFHPNTQVGLVTRYEAYLSHADEKIGIOV---RVFDED--PGPFOHV--- 378
DB 173 WK-----TSDQMDGLTTN-----GVLVHPRNGFTEDSKPGIMREISVC 212
QY 379 -----MDQISSCTQKEKLLPEVDY-----LYERSRHVNTPKHRAVL 414
DB 213 GNVSFLRETRSQAQGRKMV-EIETNOLQDGLIDLCGATLLMRTAEGLSHTPYVKHLEAL 271
QY 415 VTSLNAGYAENLKSMEWEPSTGEIIGVHPQSGYQOTEKKMHNGKALAEYLLSLTD 474
DB 272 RQEINARQ-----CPVGF-----NTLAFPSMKRKRDVVD 301
QY 475 NLVTSAMSTFGYVAGLGLKLPWILYRPENRTTPDPSCGRAMSMEPCFHSPEFYDCAKT 534
DB 302 EKQPMVYLLNCGHV---HGYNHW--GNKEERDQKDRCPMCRSVGP--YVPLMLGCEAGF 353
QY 535 GIDTG 539
DB 354 YVDAG 358

RESULT 8
US-09-815-242-12623
; Sequence 12623, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12623
; LENGTH: 876
; TYPE: PRT
```

ORGANISM: Staphylococcus aureus
US-09-815-242-12623

Query Match
Best Local Similarity 20.2%; Score 86.5; DB 10; Length 876;
Matches 95; Conservative 82; Mismatches 151; Indels 143; Gaps 28;

QY 84 GVPFVNTNSDKLGLGSLAGFDEDSCLSR-YOSVHYRKPSYPKSSYLLS-----K 135
DB 213 GVEHLEIDETSGKLKEKLONS--KEDALMSKELATINVDSPIEVLEDTIMTHOEOQEK 270
QY 136 LRNTEKLHKRCGPGETSEYKALKOLDQ-----EHIDGGECKYVWI 177
DB 271 IELFKKL-----EFKQLADIDOSASVEDALEKTFELETSDNDVDFSLKBAVILHF 321
QY 178 SFSGLGNILSLASFVALLTDRLVLDKRGKMDLFCPEPLGSMILLPDPFTDQ-- 235
DB 322 ELDC-GNYLRN--NLKFSLFTGEKHIVINADJINN-YAEL--VSW--LENPTKKVY 371
QY 236 FD-----GLNDSRCYGYMKNQVYIDFEGTSLHYLHVHYGDHDKMFCEG 284
DB 372 YDAKTTYASHRLGIDIONI-SFDMLASYIIDPSRTISDVO-SVYSLYGQS---FYKD 425
QY 285 DQTEIGKVPMLIVKTDNYFVPSLML-----PGFDELNKLFPQKATYFHHGLRYLF 336
DB 426 DVSTYGGKGFKEVPEDVLIPIYASITDAIFYAKPNMDKOLEEY----- 469
QY 337 HPTNQVGLVTRYIE-ATLSHADEKIGIQVRYFDEDPGPFQHVMDQISSCTQKEKLPB- 394
DB 470 --NOVELLADLEPLAKILSEMEIGIFTVDHLE-----EMEKEIOEK 511
QY 395 VDLVERSRL-----VNTPKRKAV-----LVTSINAGYA---ENLKSMTWEP 434
DB 512 LDVLI-RNHHVAGDEINSPKOLGVLEFLOLPVIKRTGTSTAVDVLEQJGHEHP 570
QY 435 TSTGEIIGVHQPQ-----EGYQOT---EKKMHN--GKALAEMYLSLTD 474
DB 571 I-IDVILEYRLSKLQSYVEGLQKVISDDQRIHFRNQTLAGOTRLSSVD 620

RESULT 9
US-09-843-905A-2
Sequence 2, Application US/09843905A
Patent No. US2002016863A1
GENERAL INFORMATION:
APPLICANT: Bird, Timothy A.
APPLICANT: Cosman, David J.
TITLE OF INVENTION: HUMAN PELLINO POLYPEPTIDES
FILE REFERENCE: 2990-A
CURRENT APPLICATION NUMBER: US/09/843,905A
CURRENT FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/200,198
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 418
TYPE: PRT
ORGANISM: Mus musculus
US-09-843-905A-2

Query Match
Best Local Similarity 2.9%; Score 86; DB 9; Length 418;
Matches 90; Conservative 47; Mismatches 160; Indels 188; Gaps 22;

QY 99 GGLASGF-----DEDSCLSRYOSVHYRKPSYPKSSYLLSKLRNTEKLHKRCGPSTE 151
DB 18 GELIYLVGNGSLPNDGRGRKRSRFLFKRKANGVKPST-----VHIACTP--- 63
QY 152 SYKKAQKOLDQEHIDGDECKYVWISFGAGNRLSLASFVALLTDRLVLDKRGKMD 211
DB 64 QAAKAISNKDHSIS-----YLSRAQIVVVEVYFTHDS 95

QY 212 D-DLFCPEPLGNSWLLPLDFPMTDQFDGLNDESSRCYGYMKNQVYIDT---EGTSLSHLYL 267
DB 96 NTDME---QIGRSTESPIDFVVTDFVPGSQSNS-----DTQSVGSTITSRFAC 139
QY 268 HLVDHYGDHDKMFCEGDTFYGKVPMLIVKTDNYFVPSLMLIPGFDDELNKLFPQKATV 327
DB 140 RLI-----CERSPPFYARI-----YAAFGDSSKNIFLEGKAAK 172
QY 328 FHHLGRYLFHPTNQVGLVTRYIEAYLSHADEKIGIQV---RVDEDD--PGPQHV--- 378
DB 173 WK-----TSDGMDGLTNN-----GVLVHHPNGFTESKPGIMREISYC 212
QY 379 -----MDIISCTQKEKLPEDVT-----LYERSRHVNTPKRKAVL 414
DB 213 GNVFSLRSTRSQQRGKAV-ELIETNOLDGSLIDLCGATILMRTAGLSHTPTVAKHLEAL 271
QY 415 VTSLNAGYAENLKSMTWEPYSTGTEIGVHQPQSGEYQOTERKMHNKALAEMLYSLTD 474
DB 272 ROEINAAAPQ-----CPVGF-----NTLAFPSMKRRDQVD 301
QY 475 NLVTSAMSTFGVYVAGGLGKLPWILLYREPNTKTPPSCGRAMSMERCFHSPPFYDCKAKT 534
DB 302 EKQPVVYLLNCGHV---HGYNHW--GNKEERDQKDRCPMCRSVGP--VYPLMLGCEAGF 353
QY 535 GIDTG 539
DB 354 YVDAG 358

RESULT 10
US-09-862-810-4
Sequence 4, Application US/09862810
Patent No. US20020034515A1
GENERAL INFORMATION:
APPLICANT: Mobley, Harry L.T.
APPLICANT: Li, Xin
TITLE OF INVENTION: PROTEUS MIRABILIS-BASED VACCINE
FILE REFERENCE: University of Maryland
CURRENT APPLICATION NUMBER: US/09/862,810
CURRENT FILING DATE: 2001-05-22
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 335
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: PapG fimbrial
US-09-862-810-4

Query Match
Best Local Similarity 2.8%; Score 85.5; DB 10; Length 335;
Matches 46; Conservative 25; Mismatches 70; Indels 75; Gaps 11;

QY 250 MYKNOYIDTEGTLSHLYLHVHDYG-----DHDKMFCEGDTFYGKVPMLIVKTDNFFV 304
DB 85 VVPKHYVTNEG--FNIFLDVQSKYGSMMENENDKDEY-----FVNGYEW----- 127
QY 305 PSLW-----LIPGFDDELNKLFPQKATVFNHL-----GRYLFHPTNQVGLVTRY 349
DB 128 -DTWTNNGARICFYFGNMKOLNNKF--NDLVFRVLLPYDLPKGYHNF-PVRYIRIGIOHHY 183
QY 350 YEAYLSHADEKIGIQVRYFDEDPGPFQHVMDQISSCTQKEKLPEDVTLY-----E 400
DB 184 YDLMOQH-----YKMPYDOI-----KQLPATNTLMLSFQVNGGQ 218
QY 401 RSRHVNTPKRKAVLVYTSLNAGYAENLKSMTWEPYS 436
DB 219 PSTQVLNIDHGSIVIDIRANGINIASOTLSIYCDVPVS 254

```
RESULT 11
US-09-815-242-13431
; Sequence 13431, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13431

Query Match      2.8%; Score 85.5; DB 10; Length 502:
Best Local Similarity 18.5%; Pred. No. 8.7; 176; Indels 151; Gaps 29;
Matches 96; Conservative 95; Mismatches 151; Gaps 29;

QY 58 FSMIFHQ-----HPSDS--NRINGFAEARVLDAGFPPN-----VTNINSDKLGLLAS 104
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 36 FTQIFPQAGWVHNANNEIMNSVQSIAGNFISGVKPNQIEAIGITNORETVV----- 89
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 105 GFDEDSCLSRVQSVHYKRPSPYKPSYLSKLRLN---YEKLHRCGPGTESYKALK-QL 160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 90 -WDKRTGTPYNAIWQOS---RQTAPLAEOQLKSOGYVEKFEKTEKGLIIDAVFSATKVRW 144
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 DQEHIDG-----DSEC-----KYVWISFSGLGN--RIISLASVFLYAL-----LTDRV 202
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 145 ILDHVEGAGQERAEKELLFGTITDITWLVKMLTDGAHVVDYSNAARTMYLNIEKLWDEIT 204
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 203 L-----LVDRGDMDDL--CEPFLGMSWLPFLDFPMIDQ-----FDGLNQOESSRCYG 248
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 205 LEILNIPKALIDPEVNSNSBYKTPAFHYGGEVPIISGAGQQAALRGQLAFEPG---- 260
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 249 VVAKQVQVDTESLTLVLYLHLVHD-----YGDHDKMFECBGDDQFTI--GKVP 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 261 -VWKN---TYGTGSIPTIMNGEEMQLSENLLTTIGYINGKVVYALEGSIFIGSAIQ 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 294 WL-----LYKT-----DNVFPVPSLWLPFGPDELKLPKQATYFHHGRLFHPT 339
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 316 WLROGLRWENSPSESEKARSDSHNDDEYIVVAFGLGAPYNNQNA-----R 362
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 340 NOVWGLVTR--YYEAYLSHADEKIGIOVRVEDDEGPFQHVWMDQISSCTQKKEKLLPEVD- 396
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 363 GSVFGL-TRGTSKEDFIKATQSIAYQVR-----DIIDTQVQVDTQTAIQVLAKDVG 411
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 397 -----TLVRSRHVNTPKKAVLVTSINAGYANLKSMTWEYPTSTGE 439
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 412 GAAMNFMQFQADILGIDIRAKNLETTALGAFLAGLSVGYWMDLDEL--KILNFTGE 469
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 440 IIGVHQPQSGEGYQVQTEKKMHNG--KALAEWILLSLTDN 475
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 470 LF---EFSKN--ESRKEQLYGKWKRAKATOVFAEVD 502
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-10-028-072-234
; Sequence 234, Application US/10028072
; Publication No. US2003004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gunney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
```


PRIOR FILING DATE:	1997-10-27
PRIOR APPLICATION NUMBER:	60/0633229
PRIOR FILING DATE:	1997-10-27
PRIOR APPLICATION NUMBER:	60/0635500
PRIOR FILING DATE:	1997-10-28
PRIOR APPLICATION NUMBER:	60/0635651
PRIOR FILING DATE:	1997-10-28
PRIOR APPLICATION NUMBER:	60/063704
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/0637378
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/0637555
PRIOR FILING DATE:	1997-10-17
PRIOR APPLICATION NUMBER:	60/065846
PRIOR FILING DATE:	1997-11-17
PRIOR APPLICATION NUMBER:	60/066364
PRIOR FILING DATE:	1997-11-07
PRIOR APPLICATION NUMBER:	60/065186
PRIOR FILING DATE:	1997-11-12
PRIOR APPLICATION NUMBER:	60/065846
PRIOR FILING DATE:	1997-11-17
PRIOR APPLICATION NUMBER:	60/066364
PRIOR FILING DATE:	1997-11-21
PRIOR APPLICATION NUMBER:	60/065186
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/066511
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/066770
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/069212
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069278
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069334
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/069694
PRIOR FILING DATE:	1997-12-16
PRIOR APPLICATION NUMBER:	60/072320
PRIOR FILING DATE:	1998-01-23
PRIOR APPLICATION NUMBER:	60/073612
PRIOR FILING DATE:	1998-02-04
PRIOR APPLICATION NUMBER:	60/074086
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/074092
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/077791
PRIOR FILING DATE:	1998-03-12
PRIOR APPLICATION NUMBER:	60/078910
PRIOR FILING DATE:	1998-03-20
PRIOR APPLICATION NUMBER:	60/079294
PRIOR FILING DATE:	1998-03-25
PRIOR APPLICATION NUMBER:	60/079663
PRIOR FILING DATE:	1998-02-27
PRIOR APPLICATION NUMBER:	60/079728
PRIOR FILING DATE:	1998-03-27
PRIOR APPLICATION NUMBER:	60/080155
PRIOR FILING DATE:	1998-03-31
PRIOR APPLICATION NUMBER:	60/081203
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081229
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081655
PRIOR FILING DATE:	1998-04-14
PRIOR APPLICATION NUMBER:	60/081817
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/081818
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/082999
PRIOR FILING DATE:	1998-04-24

PRIOR APPLICATION NUMBER: 60/083322	PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545	PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600	PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627	PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637	PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149	PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323	PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338	PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085339	PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579	PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697	PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704	PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086414	PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086430	PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087106	PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/088026	PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088730	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088741	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/089522	PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089529	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089597	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090349	PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090538	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090634	PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863	PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360	PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519	PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091982	PRIOR FILING DATE: 1998-07-07

```

Query Match          2.8%; Score 85; DB 9; Length 1160;
Best Local Similarity 19.3%; Pred. No. 36;
Matches 106; Conservative 70; Mismatches 187; Indels 186; Gaps 30.

QY      89 VTN--INSDKLLGLLACGF---DEDSCLSIRQSVAHYRRKPSYKSSY----- 131
        |||..|||..|||..|||..|||..|||..|||..|||..|||..|||..|||..|||
IDb     50 ITNPLDLD---IVASSFLKSDNNRIGTGYKKTKIYRK--YKDDSYDEVAOPAMLCG 101

```

```
QY 132 -----LISKLNRYEKLHKRCGPGETSYKK-----ALKOLDQ----- 162
Db 102 LCPVLQAEVGVILHLKKNFATRPYTIHPHGVFEKESSEGLYPDGSSGPKADDSVPPG 161
QY 163 -EHI-----DGGCECKYVWISFSGLNRIISLASVFLYALLTDVRLVLD--- 206
Db 162 GSHLYNWTIPESHAPTDAPAC--LTWYIHSHV-DAPRDIAITGLIGPLITCKRGALDGN 218
QY 207 --RGKMDM-----LFCPEFLGMSWLP-----LDFPMTOQFDGLNDESSRCY---GYM 250
Db 219 PQQRQDVHDHDFLLFSYVDENLSMHLNENIATYCSDPASVDKEDETFQESNMHAINGEV 278
QY 251 VKNQVIDTEGTLSHLYL-----HL-----VHDYGDHDKMFCEGDQTFIGK--- 291
Db 279 F-----GNLPFLNMQAOKRVAMHLFGMGNEIDVHTAFPHGQMLTTRGHHTDVANIFP 330
QY 292 -----VPMILVKTNDVYVPSLMLI-----PGFDELNKLFPQKATVPHHLGRYLEHP 338
Db 331 ATFTVTAEMVPM-----EPGTWLISCQVNSHFRDGMQALYKVKSC-----SNAPP 374
QY 339 TNOVGLVTRY-Y-EAY-----LSHADEKIGIOVRFPEDPGP-----FOHVMQISS 384
Db 375 VDLITGKVRQYFIEAHNLIQMDYGRMGH-DGSTGKNLR---EPGISIDKFFQKSSSRIGG 429
QY 385 CTQKEKLLPEVDTLVRSRHYNTPKHKAVALVTSLNAGYAENLKSMTWEYPTSTGEITGVH 444
Db 430 TYMKVRYEAFQDETFQEKHMLEEDRHILGPIRAEVGDTIQVVFYNNASQPSFM----- 485
QY 445 QPSQEGYQOT-EKKMHNGKALAEMLYLLSLDNLVTSMSFTGYVAAOGLGKLPWILYRPE 503
Db 486 QPHGVFEKDYEGTYVNDG-----SSYPGLVAKPEKVTY-----RWTVPPHA 528
QY 504 NRTTPDPSC 512
Db 529 GPTAODPAC 537

RESULT 13
US-10-121-049-234
; Sequence 234, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-234
Query Match 2.8%; Score 85; DB 9; Length 1160;
```

```
Best Local Similarity 19.3%; Pred. No. 36;
Matches 106; Conservative 70; Mismatches 187; Indels 186; Gaps 30;
QY 89 VTN--INSPKLLGGLLASEF---DEDSCLSRQSVHYHRRPSYKPSY----- 131
Db 50 ITNQPLDSD-----IYASFLLSKDKNRIGGYTKKITYKE--YKDSYSTDEVAQPAWLGF 101
QY 132 -----LISKLNRYEKLHKRCGPGETSYKK-----ALKOLDQ----- 162
Db 102 LCPVLQAEVGVILHLKKNFATRPYTIHPHGVFEKESSEGLYPDGSSGPKADDSVPPG 161
QY 163 -EHI-----DGGCECKYVWISFSGLNRIISLASVFLYALLTDVRLVLD--- 206
Db 162 GSHLYNWTIPESHAPTDAPAC--LTWYIHSHV-DAPRDIAITGLIGPLITCKRGALDGN 218
QY 207 --RGKMDM-----LFCPEFLGMSWLP-----LDFPMTOQFDGLNDESSRCY---GYM 250
Db 219 PQQRQDVHDHDFLLFSYVDENLSMHLNENIATYCSDPASVDKEDETFQESNMHAINGEV 278
QY 251 VKNQVIDTEGTLSHLYL-----HL-----VHDYGDHDKMFCEGDQTFIGK--- 291
Db 279 F-----GNLPFLNMQAOKRVAMHLFGMGNEIDVHTAFPHGQMLTTRGHHTDVANIFP 330
QY 292 -----VPMILVKTNDVYVPSLMLI-----PGFDELNKLFPQKATVPHHLGRYLEHP 338
Db 331 ATFTVTAEMVPM-----EPGTWLISCQVNSHFRDGMQALYKVKSC-----SNAPP 374
QY 339 TNOVGLVTRY-Y-EAY-----LSHADEKIGIOVRFPEDPGP-----FOHVMQISS 384
Db 375 VDLITGKVRQYFIEAHNLIQMDYGRMGH-DGSTGKNLR---EPGISIDKFFQKSSSRIGG 429
QY 385 CTQKEKLLPEVDTLVRSRHYNTPKHKAVALVTSLNAGYAENLKSMTWEYPTSTGEITGVH 444
Db 430 TYMKVRYEAFQDETFQEKHMLEEDRHILGPIRAEVGDTIQVVFYNNASQPSFM----- 485
QY 445 QPSQEGYQOT-EKKMHNGKALAEMLYLLSLDNLVTSMSFTGYVAAOGLGKLPWILYRPE 503
Db 486 QPHGVFEKDYEGTYVNDG-----SSYPGLVAKPEKVTY-----RWTVPPHA 528
QY 504 NRTTPDPSC 512
Db 529 GPTAODPAC 537

RESULT 14
US-10-123-904-234
; Sequence 234, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:26:28 ; Search time 40 Seconds
(without alignments)
1858.846 Million cell updates/sec

Title: US-10-037-311a-1
Perfect score: 3004
Sequence: 1 MQNSTYRRSSPIRTTGGSS.....GTLPHVRHCEIDISWCLIV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1687.5	56.2	509	23	ABR90986 Herbicidally activ
2	1550	51.6	535	21	AA50465 Arabidopsis thalia
3	1530	50.9	500	21	AA50466 Arabidopsis thalia
4	1354.5	45.1	426	21	AA50467 Arabidopsis thalia
5	1124.5	37.4	562	23	ABR91595 Herbicidally activ
6	651	21.7	254	21	AA608487 Arabidopsis thalia
7	631	21.0	219	21	AA608488 Arabidopsis thalia
8	457	15.2	146	21	AA608489 Arabidopsis thalia
9	410	3.7	658	19	AA98873 H. pylori GHPD 173
10	103.5	3.4	409	22	ABR52717 Escherichia coli p

11	99.5	3.3	439	22	ABR68433
12	99	3.3	769	23	AAU93188
13	99	3.3	814	23	AA017586
14	99	3.3	818	22	AB119945
15	98	3.3	745	22	AA67331
16	98	3.3	864	22	AA692894
17	97	3.2	446	18	AAW20095
18	97	3.2	594	18	AAW21020
19	97	3.2	2044	23	ABR47302
20	97	3.2	2506	22	ABG07191
21	97	3.2	2506	22	ABG30064
22	96.5	3.2	1320	22	ABR66149
23	96	3.2	520	23	ABP60956
24	96	3.2	524	23	ABP60954
25	96	3.2	528	23	ABP60955
26	96	3.2	1240	23	ABP29556
27	95	3.2	1802	21	AA18217
28	95	3.2	2932	18	AAW19675
29	93	3.1	723	23	ABP40215
30	93	3.1	1817	21	AB18301
31	93	3.1	4134	20	AA731946
32	92.5	3.1	407	23	ABR06071
33	92.5	3.1	993	17	AA695268
34	92.5	3.1	993	20	AA706666
35	92	3.1	496	21	AA644216
36	92	3.1	496	18	AAW38188
37	92	3.1	808	21	ABR28602
38	91.5	3.0	1295	22	ABR64605
39	91	3.0	554	21	AA791659
40	91	3.0	554	22	AA694489
41	91	3.0	419	23	ABR77728
42	90.5	3.0	419	23	ABR07921
43	90.5	3.0	452	21	AA607112
44	90.5	3.0	502	22	AA638067
45	90.5	3.0	511	21	AA607111

ALIGNMENTS

RESULT 1
ID ABR90986 standard; Protein: 509 AA.
XX
AC ABR90986;
XX
DE 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 197.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN MO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001MO-EP09892.
XX
PR 28-AUG-2001; 2001MO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI: 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX

Drosophila melanog
Arabidopsis transc
M catarrhalis MCAL
Moraxella catarrha
Human neuron prog
Human protein sequ
H. pylori cytoplas
Listeria monocytog
Novel human diagn
Drosophila melanog
Mus musculus thior
Mus musculus thior
Mus musculus thior
Streptococcus poly
Plasmodium falcipa
ATM mutant 2467del
Staphylococcus epi
Plasmodium falcipa
Plasmodium falcipa
Human NS protein s
pre-nisin modifica
Nisin B of lactoba
Arabidopsis thalia
Arabidopsis SCARC
Protein encoded by
Drosophila melanog
Human secreted pro
Human protein sequ
Murine pellino-2 p
Arabidopsis thalia
Streptococcus pneu
Arabidopsis thalia

PS Claim 5; SEQ ID NO 197; 261bp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 509 AA;

Query Match 56.2%; Score 1687.5; DB 23; Length 509;
Best Local Similarity 62.9%; Pred. No. 5e-160;
Matches 326; Conservative 68; Mismatches 93; Indels 31; Gaps 8;

QY 54 VLVAFSMTF-HQHPDSNRINGFAFARVLDAGVPNPNTNINS---DKLGLGLASGFDE 108
Db LLSFSNIFKHQ-----LLGATL--NVGSKDSVVKPRDLGLGLTADPDE 45
:
QY 109 DSCSRYSQSVHRRKRPKPPSSYLISKLANYEKLHRCGPTESTKALKQDDEHID-- 166
Db DSCSRYSQSVHRRKRPKPPSSYLISKLANYEKLHRCGPTESTKALKQDDEHID-- 166
:
QY 167 --GDGECKYVWVISEFGNGNRIISLASFYALITDRVLTVDRCMDLDLCEPPLGMSW 224
Db DSDGECKYIVWVAVYGGNRIITLASVFLALITERRIITVDQRDIDSLCEPPLGTSW 165
:
QY 225 LPLDFFMTDQFDGIDNDSRCYGYWKNQYIDTEGLSHLYHLVHDYGDHKKMFCEG 284
Db LPLDFFMTDQFDGIDNDSRCYGYWKNQYIDTEGLSHLYHLVHDYGDHKKMFCEG 284
:
QY 285 DOTFGKYPMLIVTDNNEFVPSIMLIPGFDEINKLFQKATVPHNLGRVLEPHNQWVG 344
Db LPLDFFMTDQFDGIDNDSRCYGYWKNQYIDTEGLSHLYHLVHDYGDHKKMFCEG 285
:
QY 226 DQSLVDKYPMLIVTNSLTFPSLWNPSPFQTELKLPQOKIVFHPLARYLPHPTNQWVG 285
Db DQSLVDKYPMLIVTNSLTFPSLWNPSPFQTELKLPQOKIVFHPLARYLPHPTNQWVG 285
:
QY 345 LVTRYEAYLSHADEKIGIQVRVEDDEDPFOHVMQDQSSCTQKFKLPEVDTLVERSRH 404
Db LVTRYEAYLSHADEKIGIQVRVEDDEDPFOHVMQDQSSCTQKFKLPEVDTLVERSRH 404
:
QY 286 MVTNSYNAYLSRADEILIGIQRVFSRQTKYFOHVMQDQVACTQPKILPEF-AADEEAY 344
Db MVTNSYNAYLSRADEILIGIQRVFSRQTKYFOHVMQDQVACTQPKILPEF-AADEEAY 344
:
QY 405 VNT---PRHKAIVLTSNAGYAENLKSWMYEPYSTGEIIGVHOPSDQGYOOTEKMHNG 461
Db VNT---PRHKAIVLTSNAGYAENLKSWMYEPYSTGEIIGVHOPSDQGYOOTEKMHNG 461
:
QY 345 TMTSNPSKILKAVLTSNLPREXSNLKKMYMEHPPTTGDIVEYQSRERFOOTKKLHDQ 404
Db TMTSNPSKILKAVLTSNLPREXSNLKKMYMEHPPTTGDIVEYQSRERFOOTKKLHDQ 404
:
QY 462 KALAEMLYLTLDNLVTSANSTFGVVAOGLGKLPWILYRENDRTTPRSCGRAMSEPC 521
Db KALAEMLYLTLDNLVTSANSTFGVVAOGLGKLPWILYRENDRTTPRSCGRAMSEPC 521
:
QY 405 KALAEMLYLTLDNLVTSANSTFGVVAOGLGKLPWILYRENDRTTPRSCGRAMSEPC 464
Db KALAEMLYLTLDNLVTSANSTFGVVAOGLGKLPWILYRENDRTTPRSCGRAMSEPC 464
:
QY 522 FHSPPEYDCKAKGTGIDTGLVPHVHNCEDI--SWGKLTV 558
Db FHSPPEYDCKAKGTGIDTGLVPHVHNCEDI--SWGKLTV 558
:
QY 465 FLTPPVHGCCEAKKGINTAKIVPFVRHCEDLRHXYGLTV 502
Db FLTPPVHGCCEAKKGINTAKIVPFVRHCEDLRHXYGLTV 502
:

RESULT 2
ID AAG50465
ID AAG50465 standard; Protein: 535 AA.
AC AAG50465;
XX
DE 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63958.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX

PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

DB 429 DKVHDKALAEMLILSLDNTIVASSRSTFGYVAYSIGLKPWLLPLDNDKAPDPCCVR 488
OY 515 AMSMEPCFHSPEFYDCKAKT-GIDGTLLPHVRHGEDISWGILK 557
DB 489 SISMPCFLTFPTGCGECPDAKGTESGKVPFVRYCEDI-WGILK 531

RESULT 3
ID AAG50466
AAG50466 standard; Protein: 500 AA.

AC AAG50466;
XX
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63959.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.9%; Score 1530; DB 21; Length 500;
Best Local Similarity 58.5%; Pred. No. 3.2e-144;
Matches 299; Conservative 72; Mismatches 118; Indels 22; Gaps 9;

OY 54 VIVASMTFHQHSNDRIMGFAEAVDAGFPPVNTINSKLLGGLASGDFDSDCLS 113
DB 1 MLTSSNNF-----NNKL--FA-ATIDSESTP-----GRDLILGGLTADFDGSCSLS 46
OY 114 RYOSVH-YRKSPYKPPSSYLISKLNRYEKLHKRCGPTGESYKALKQLDQDEH----IDOD 168

DB 47 RYHKTFLYRKPSPELVSKLSYEMLHKRCBPGRKAYEATKHLSHDENVAKSD 106
OY 169 GECKYVWISFSGLCNRLISLASVFLYALLDRVLLVDRGKMDLFCBEPFLGMSMLPL 228
DB 107 GECRYVWVLADYGLCNRLITLASVFLYALLDRILVDRKIDGLDCEPFGTSMPLPL 166
OY 229 DEPRMDQFDGLNQESSRCGYVWKNQVIDTECTLSHLYLHVHVDGDDHKMFCEGGDOTF 288
DB 167 DEPLMKYADGTHKGYSRGGLTLENHSTISFPPLHYMHNLDSDSDSKMFCCKDOSL 226
OY 289 IGVPMVLVKTNDVFPVPSLMLTGFDDLELNKLPQKATVFHHLGRTLEPHTNQVGLVTR 348
DB 227 IDKVPMLIFRAVYFVPSLWMPFTQFELTKLFPQKEYVFHHLGRTLEPHTNQVMDIYTK 286
OY 349 YVEAYLSHADEKIGTIOVRFEDEPGPFQHWMDQISSCTOKEKLEPVDLVERSNNV-NT 407
DB 287 YYHDLISKADERLGTIOIRVFROGGVYOHVMDQVISCQDREKILBELATQEESSKNISNI 346
OY 408 PKHKAIVLTSLNAGYAENLKSWMYEPTSTGELIGVHOPSQGYOOTEKMHNGKALAM 467
DB 347 PSKAIVLTSLSPEYSKLENNFSEKANTGELIKVYOPSGERYQOTDKKVDQKALAM 406
OY 468 YLSTLTDNLVTSAMSTFGYVAAOGLGLKPMILYRPENRTTDPSCGRAMBPCFHSPPF 527
DB 407 YLSTLTDNLVTSAMSTFGYVAAOGLGLKPMILYRPENRTTDPSCGRAMBPCFHSPPF 466
OY 528 YDCKAKT-GIDTGLVPHVRCEDISWGLK 557
DB 467 HCCEPDAMCTESGKVVPFYRCEDI-WGLKL 496

RESULT 4
AAG50467
ID AAG50467 standard; Protein; 426 AA.
AC AAG50467;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63960.
DE
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
EP
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134761.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140821.
PR 29-JUN-1999; 99US-0140923.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0143390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144083.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145215.
PR 23-JUL-1999; 99US-0145215.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145275.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157713.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

```
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.1%; Score 1354.5; DB 21; Length 426;
Best Local Similarity 60.7%; Pred. No. 1e-126;
Matches 256; Conservative 60; Mismatches 99; Indels 7; Gaps 4;

OY 142 LHKRCGPCTESYKKAALKOLODEH---IDGDECKYVWISSGHNRLSLASVFLYL 197
DB 2 LHKRCGPCTKAYKEATKHLSDENYNAKSDGECRYVWLADYGLGNLLTLASFYAL 61

OY 198 LTRVLLVDKGMKMDLFCERFLGSMWLLPLDFPMTDOFDGLNDESSRCYGYMKNOYID 257
DB 62 LTRRIILVDNKRKIDGLLCEPFPCTSMLLPLDFPLMKYADGYHKGYSRCYGTMLENHSIN 121

OY 258 TEGTSLHYLVLHYHGDHDKMFCEGDOTFGKVPMLIVKTDNFVFSIMLIPGDEDL 317
DB 122 STSEPHLYMNLHDSRSDKMFECQKDSLDKVPMLIFRANVYFVSLMNPFFQEL 181

OY 318 NKLFQKATVFHNLGRYLFHPTNQVGLVTRYEAYLSHADEKIGIOYRVFDEDEGPFQ 377
DB 182 TKLFQKQETVHHLCRYLFHKNQVMDIVTKYHDLKADERLQIRVFDQGGYIOH 241

OY 378 VMDQISSTQKELLPEVDLTVERSRHV-NPKRAVLVTSINAGYAENLKSWMYEYPTS 436
DB 242 VMDQVISTQREKLPELATQDESKVNISNIPKRAVLVTSISPEYSKLEMFSEKRAM 301

OY 437 TGEITGVHQPQOEQYQOEKKMHNGKALAEWYLLSTLNLVTSAMSTGCVYVAGGLK 496
DB 302 TGEITKVOPSGERYQYQDKVHODKALAEWYLLSTLNLVTSAMSTGCVYVAGGLK 361

OY 497 WILYPRNRTTDPSCSGRAMSWPCFHSPPFYDCKAKT-GIDTGLVPHVRHCEDISGL 555
DB 362 WILYPRNKNKAPDPPCVASTJSMRPGFLTPPHGCEPDAMGTESGKVPFVRYCEDI-WGL 420

OY 556 KL 557
DB 421 KL 422

RESULT 5
ABB91595
ID ABB91595 standard; Protein; 562 AA.
AC ABB91595;
XX
XX 31-MAY-2002 (first entry)
DE Herbicidally active polypeptide SEQ ID NO 806.
KW Herbicidal; plant; agriculture; herbicide.
OS Arabidopsis thaliana.
PN W0200210210-A2.
PD 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP09892.
XX
```

```
PR 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidner M;
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 806; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 562 AA:
XX
XX Query Match 37.4%; Score 1124.5; DB 23; Length 562;
XX Best Local Similarity 45.2%; Pred. No. 2e-103;
XX Matches 242; Conservative 83; Mismatches 163; Indels 47; Gaps 13;

OY 15 TTTGSGKSVNFEELQMYLSSGTMKLTFTTCTLVFSVLAVAFSMIFQHPSDSNRIMG 74
DB 56 TRTGS-----SELNMMKRPSSLSKMTGMLLAVLVASVMSLSVLRDPPSD-DVLETT 108

OY 75 FAEARVLDAGVFPNVTNINSDKILGL-----LASGDESDCLSRQSVHYHRKP 123
DB 109 EASRVSLSRLH--QATESD---GLSEKKAQLNINLVPFDEKSCLSRYEASLYRKE 162

OY 124 SPYKRSYLSLKLRYEKLHKRCGPCTESYKKAALQOLO-ELIIDD-GECKVYVVISFG 181
DB 163 SPFKQSYLDVRLQRYEDLHRCGPFTRSYNLTLDKLSGSDSEVSGCRVYIWLNSNG 222

OY 182 -LGNRLSLASVFLAALLTRVLLVDKGMKMDLFCERFLGSMWLLPLDFPMTDOFDGLN 240
DB 223 DLGNMMLSLASFLLALTLNRLVLGLVDADLCEPFPNTTWLPLPEFPLANSF---N 279

OY 241 QESSRCYGYMKNOYIDTEGTLSHLYLHLVNDYGDHDKMFCEGDOTFGKVPMLIVKTD 300
DB 280 EQS-----LLRNSGNPMVAVRYRHVYRDSQOKLFCEDSOVLLLETPMLILKAD 328

OY 301 NYFVPSLMLIPGDEDLNKLFPQKATVFHNLGRYLFHPTNQVGLVTRYEAYLSHADEK 360
DB 329 SFELPSLFSVSSFKOELMLFPEKDTAFHLSQYLFHPTNVWGLITRYNNYLLAKAQR 388

OY 361 IGIOYRVFDEDEGPFQOHVMDQISSTQKELLPEVDLTVER---SRHVTPKRAVLVTS 417
DB 389 IGIIYIGVSESGNEQFQHLIDQILACGTRKHLPEVDK--QRNLPSOVLNKRKSAVPISS 446

OY 418 LNAGYAENLKSWMYEPSTGEIIGVHQPQOEQYQOEKKMHNGKALAEWYLLSTLNL- 476
DB 447 SSPGYFKSIRDYWMENPVTYGEIISVHKPSYDQXKTPRNMSKRAMAEIYLLSCDALV 506

OY 477 VTSAMSTGCVYVAGGLKLPMLYPRNRTTDPSCSGRAMSWPCFHSPPFYDCK 531
DB 507 VTGLMWSLVEVAHGLGGLKLPWLVNKAENGTAHEPYCVAKARSIPCSQATLTFHGCK 561

RESULT 6
AAG08487
ID AAG08487 standard; Protein; 254 AA.
AC AAG08487;
XX
```

```
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 6045.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0128845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
```

```
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 21.7%; Score 651; DB 21; Length 254;

Best Local Similarity 53.4%; Pred. No. 1.8e-56;

Matches 132; Conservative 36; Mismatches 59; Indels 20; Gaps 6;

```
OY 41 LTRPTTCLIVSVAVSMIFHQHPDSNRIMGFAPARVLDAGVFPVNTNNSKILGG 100
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 23 ITIVSTLILSLVMLSSNNF-----NNKL--FA-ATINDESETP-----GRLTGG 67

OY 101 LLASGFEDSDCLSRXYOVH-YRKPSYPKPSYLSKLNRYEKLHNRGPGTESYRKALQ 159
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 68 LLTTFDEGSCSLRYHKNFLYRKPSYPKPSYLSKLNRYEKLHNRGPGTESYRKALQ 127

OY 160 LDQEH-----IDGDECKKYVMAISFGGNGRIISLASVFLYALITDRLVLDGKMDLE 215
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 128 LSHDENYNAKSKDGCRRYVWLADYGLNRLTLTASVFLYALITDRIILVDMRKDISDL 187

OY 216 CEPFGMSWLPFLDPFMTDQFDGLNOESSRCYGVWKNNOVDTDEGLSHLYLHVHDYGD 275
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 CEPFGTSHLPFLDPFLKRYADGYHKYSRCYGTWLENHNSINSTSPPHLWMLHDSRD 247

OY 276 HDKMFEC 282
```

```
Db 248 SDKMFEC 254
|||||
RESULT 7
AAG08488
ID AAG08488 standard; Protein: 219 AA.
XX
AC AAG08488;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6046.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
```

[illegible]

PR	17-AUG-1999	9905-0119175
PR	18-AUG-1999	9905-0119426
PR	20-AUG-1999	9905-0119722
PR	20-AUG-1999	9905-0149723
PR	20-AUG-1999	9905-0149929
PR	23-AUG-1999	9905-0149929
PR	23-AUG-1999	9905-0149920
PR	25-AUG-1999	9905-0105666
PR	26-AUG-1999	9905-0150884
PR	27-AUG-1999	9905-0151066
PR	27-AUG-1999	9905-0151066
PR	30-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151438
PR	01-SEP-1999	9905-0151930
PR	07-SEP-1999	9905-0152363
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154179
PR	22-SEP-1999	9905-0155133
PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155569
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156396
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159637
PR	14-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160747
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160982
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

Query Match 21.08; Score 631; DB 21; Length 219;

Matches 127; Conservative 33; Mismatches 54; Indels 20; Gaps 6;

[illegible]

QY 169 GECKYVWISFSGICGNRIILSLASVFLYALTDRLVLDVGRKMDMLFCEPFLGSMLLPL 228
|||:||||: ||||:|||||:||||: ||: || |||| | |||||
Db 106 GECRVVWLAQYGLGNRIILTLASVFLYALLDRILVDRNRKDIDSLCEPFGTSMLLPL 165
QY 229 DFPMTDOFDGLNOESSRCYGVKQNVIDTRECTLSHLVHLVHYDGDHDKMFEC 282
|||: ||: ||||| ||: ||: ||||| ||: ||| || |||||
Db 166 DFPMLAKYADGYHKGYSRCYGTMLENHSINSTSFPHLYMHNLHDSRDSDKMFEC 219

RESULT 8
AAG08489
ID AAG08489 standard; Protein: 146 AA.
XX AAG08489;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6047.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

```
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155658.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.2%; Score 457; DB 21; Length 146;
Best Local Similarity 58.3%; Pred. No. 2.2e-37;
Matches 84; Conservative 21; Mismatches 35; Indels 4; Gaps 1.
```

```
Oy 143 HKRCGPGRSTYKALKOLDQEH---IDGCGCKTYVMISSGCLNRILISLASFVFLVALL 198
Db 3 HKRCGPGRSTYKALKOLDQEH---IDGCGCKTYVMISSGCLNRILISLASFVFLVALL 62
```

```
Oy 199 TDRVLLVDRGKMDLFCGEPFLGMSWLLPDPFMTDQFDGLNQESSRCGYMKNOVIDT 258
Db 63 IDRIILVDRNRKDISDLCEPFGTSMWLLPDLPMKRYADGYHKGYSRGYTMLENHINS 122

Oy 259 EGTLSHLYLHLVHDYGDHDKMFC 282
Db 123 TSFPHLYMHNHDSRDSKMFEC 146

RESULT 9
AAW98873
ID AAW98873 standard; Protein, 658 AA.
XX
XX AAW98873;
AC
XX
XX 31-MAR-1999 (first entry)
DT
XX
DE H. pylori GHP0 1739 protein.
XX
XX GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
XX peptic ulcer disease.
XX Helicobacter pylori.
OS
XX
XX MO9843478-A1.
PN
XX
XX 08-OCT-1998.
PD
XX
XX 01-APR-1998; 98WO-US06371.
PF
XX
XX 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX A1-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
XX DR N-PSDB; AAX14592.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
XX for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
XX
XX Claim 8; Page 1981-1984; 2054pp; English.
XX
XX This sequence represents a Helicobacter pylori GHP0 protein of the
XX invention. The polypeptides can be used for preventing or treating
XX Helicobacter infections, and gastroduodenal diseases associated with
XX these infections, including acute, chronic, and atrophic gastritis, and
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX used for the production of antibodies. The products can also be used for
XX detection and diagnosis.
XX
XX Sequence 658 AA:

Query Match 3.7%; Score 110; DB 19; Length 658;
Best Local Similarity 17.8%; Pred. No. 0.18;
Matches 102; Conservative 79; Mismatches 191; Indels 202; Gaps 23;
```

```
Oy 37 CTMKITRFTTCLIVFSYVAFSMIFHOHPSDSNRIMGFAEAVLDAGVPPNVTNINSK 96
Db 28 GQORLT-TITLLILA-----LRNHLSEVEILEKFSRKEIESYLIN--SNKGDG 74

Oy 97 ILGGILSGFEDCLSRYSVHYRKSPYKPSYLSKLNRVKKLRCGPGRESYKKA 156
Db 75 KF-RLLISESDKTLULSLDK-NRKPS--EPYKIVENFELFEKWSIENTDKLETFGK 130

Oy 157 LKOLDQEHIDGCGCKYVWIS-----FSGGNRILISLASFVFLVALLDRVLLV 205
Db 157 LKOLDQEHIDGCGCKYVWIS-----FSGGNRILISLASFVFLVALLDRVLLV 205
```



```

Db 131 LKKL-----MIWISLDKGDQDPLFESMNSKDIELTQF---DLIRNYIME 175
QY 206 DRGKMDMLFCEPFLGSMWLLPLDFPMTDOF----- 236
Db 176 TEVKKQEDFYNQYKRAMEEFEQNETLFNFRVRYLYLTIKGKIPNEKRYEAFQDQK 235
QY 237 ---DGINOESSRCYGYW-----KNQVIDTEGTLSh-----LYLHVHDYGDH 276
Db 236 GIEIEDLLKDLQKCYGFCOIAFEKEDDKDLNKALSLVNLMDVIVPLLELYSDYKD- 294
QY 277 DKMFEGSGDQTFIGKVPWLVTNDNVFVPSMLIRPF-----DDELNKLFP---- 322
Db 295 -----GVLSKQD--FIPITLYLESYICRRVAGGLGTNSLNKVPSPFTK 335
QY 323 --OKATVFHHLGRLYF-----HPTNQVWGLV-----TRYEAVLSHADEKIG 362
Db 336 HIQKDEFFKSLKNAFVCLTEKQRFNNDERFKLEFITIDYFKFKNKKFLFLELEMPDKEP 395
QY 363 IQVRVFEDEDPPOHVMQDISCTQKEKLLPEVDTLVERSHV--NTPKRAVLVTSLSA 420
Db 396 V-----DTQKCNIEHIMPQTLTPEMQRLDGENFQAIHEKYLHTIGN-----LTLTGYS 444
QY 421 GYAEN-----LKSM-----TWETPST 437
Db 445 KYSNSFOEKRDMEKGFQSSLSKLKQSLKQESFEKEIEKRASDLADWALKLTWPTLE 504
QY 438 GEIGVHPQSQEQYQOTEKKMHNKALAEWYLS 471
Db 505 AETLEEKPKKKEKKEKEKYLKREKAYDLS 538

```

RESULT 10
 ABB52717
 ID ABB52717 standard; Protein; 409 AA.

```

XX AC ABB52717;
XX DT 11-FEB-2002 (first entry)
XX DE Escherichia coli polypeptide SEQ ID NO 835.
XX KM Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
XX KM immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX KM systemic infection; non-diarrhoeal infection; septicemia;
XX KM pyelonephritis; antibiotic resistance.
XX OS Escherichia coli.
XX PN WO200166572-A2.
XX PD 13-SEP-2001.
XX PF 12-MAR-2001; 2001WO-EP03445.
XX PR 10-MAR-2000; 2000EP-0003145.
XX PR 02-FEB-2001; 2001EP-0001449.
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX DR WPI; 2001-550253/61.
XX PT A library of DNA fragments of Escherichia coli strains for the
XX PT phylogenetic determination of a given strain comprises polynucleotides of
XX PT nature B2/D+ A-
XX PS Example 6: Fig 6; 646pp; English.
XX CC The invention relates to a library of DNA fragments of Escherichia coli
XX CC strains comprising polynucleotides (AB868577-AB88729 and AB89533)
XX CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
XX CC B2/D+A-. The polynucleotides have potential antiinflammatory,

```

CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.

Sequence 409 AA;

Query Match 3.4%; Score 103.5; DB 22; Length 409;

Best Local Similarity 24.1%; Pred. No. 0.36;

Matches 53; Conservative 26; Mismatches 68; Indels 73; Gaps 11;

```

QY 389 EKLPE-----VDTLVERSHVNPKE-----HKAVLVYSLANG--YAENLKSMTVEYP 434
Db 193 EELCPSEAAVEVDLFAVSQNDADKTPSPDENRDGLVIGEGATLLELEHAKARGA 252
QY 435 TSGETIG-----VHPSQEQYQOTEKK-----MH 459
Db 253 TTYEIVFPATNCDAAHITOPQRETMOYCMQSLKIGLSAODIGYISAGTAARDGMA 312
QY 460 NGKALAEWYLSLTNDL-VTSAMSTFGYVAGLGELKPMI-----LYRPE-NRTTPD 509
Db 313 ESLATATVLY-----GDNPVVSLSKSYFGHTLGACGALAEWMSLQMMRQGFAPTLNKKPD 368
QY 510 PSCGRAMSMEPCFHSPPPYDCK-----AKGTIDTGLTV 542
Db 369 PNCG---ALDYIMHARKVDCEFLQSNMFAGGINTSITI 405

```

RESULT 11

ABB68433
 ID ABB68433 standard; Protein; 439 AA.

```

XX AC ABB68433;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 32091.
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL12536.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 32091; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and

```

CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 CC
 CC Sequence 439 AA;
 SO
 Query Match 3.3%; Score 99.5; DB 22; Length 439;
 Best Local Similarity 21.4%; Pred. No. 1;
 Matches 88; Conservative 66; Mismatches 126; Indels 131; Gaps 27;
 Oy 109 DSCLSRYQSVHNRKPPSPSSYLISLR-----NNEKHLKRCGPGESKAKMLK 158
 Db 20 ESLVFEESVFLDAPE-WLTESYLODALRKYYKQDOLITINWKPALGKG-ENYGVLT 77
 Oy 159 QLDQEHIDGDECK--YVWISFSG-----LGNRTISLASFVFLYALLTDRVL 203
 Db 78 RKKAQFTSRDSSQLGHYIVKSTFEGNEFAQNAKPYDIFNREM--IIEQVLPKOKAL 134
 Oy 204 LVDRCKMDLDFCEPF---LGNMULPLD-----FPMIDQFDGLNQESSRCYGVAKMOV 255
 Db 135 LREIG-DEQJFAETMAVDINDSALIFEDLNARGFVMPDRVLGDLQKLAR---IVLRKL 189
 Oy 256 ID---TEGTLSHLYLHVHDGDKMFCEGDDQTFIKKVPWLYKTPNY---FVPSL-- 307
 Db 190 AKMATSAVLENREHILLESY---DRGFNR-----YTONEPAFAGMLQA 232
 Oy 308 -----WLIPGED--DELNKLFPQKATVPHILGRYLFHPT-----NQWV--GLV 346
 Db 233 ATRVAVQW--GGEYKAAEIKALVP---IYMEICKRIFDISPGHINVLAHGDLWTNNVL 286
 Oy 347 TRYVAVYLSHDEKIG--IOVRVPEDEGPRQH-----VMDQISSCOKEKILPEVTL 398
 Db 287 VKY-----DKQGEPIDIVIID-----FOYTAWGSPALD-----LEYFMSS 323
 Oy 399 VERSHNVTPKHKAVALVTSLNAGYAEMLKSMWEXPTSGEITGVHPSOE 449
 Db 324 LEFDLHQN---HQEQLIVYFRHFADTLKLT--QYRSTIPSLHCHQDLQ 369
 RESULT 12
 ID AU093188 standard; Protein; 769 AA.
 XX AU093188;
 AC
 XX 02-JUL-2002 (first entry)
 DT
 XX Arabidopsis transcription factor #226.
 DE
 XX Arabidopsis transcription factor #226.
 XX Agriculture; metabolic chemical; environmental stress; drought;
 KW microbial disease resistance; herbicide resistance; seed yield;
 KW fruit yield; growth rate; leaf senescence; flower senescence.
 KW plant; transcription factor; transgenic.
 XX Arabidopsis thaliana.
 OS
 XX WO200215675-A1.
 PN
 XX 28-FEB-2002.
 PD
 XX 22-AUG-2001; 2001WO-US26189.
 PE
 XX 22-AUG-2000; 2000US-227439P.
 PR 16-NOV-2000; 2000US-0713994.
 PR 16-APR-2001; 2001US-0837944.
 XX (MEND-) MENDEL BIOTECHNOLOGY INC.
 PA (PILG/) PILGRIM M.

PA (CREE/) CREELMAN R.
 PA (DUBE/) DUBELL A J.
 PA (HEAR/) HEARD J.
 PA (JIANG/) JIANG C.
 PA (KEDD/) KEDDIE J.
 PA (ADAM/) ADAM L.
 PA (RATC/) RATCLIFF O.
 PA (REUB/) REUBER J L.
 PA (RIEC/) RIECHMANN J L.
 PA (YUGG/) YU G.
 PA (PINE/) PINEDA O.
 PA Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J,
 PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
 DR WPI: 2002-292022/33.
 DR N-PSDB: ABK65374.
 XX
 PT An isolated or recombinant polynucleotide used to produce a transgenic
 PT plant -
 PS Claim 40; Page 914-918; 941pp; English.
 XX
 CC The invention relates to 1 of 232 isolated or recombinant polynucleotides
 CC encoding an Arabidopsis thaliana transcription factor, their variants,
 CC complements, fragments, or related polynucleotide with 31% to 95%
 CC sequence identity, where the plant possesses an altered trait as compared
 CC to a wild-type or reference plant, or the plant exhibits an altered
 CC phenotype as compared to a wild-type or reference plant, or the plant
 CC exhibits ectopic expression or altered expression of one or more genes
 CC associated with a plant trait as compared to a wild plant. Also included
 CC are a transgenic plant comprising the polynucleotides, a computer
 CC readable medium having stored sequence information, and identifying a
 CC homologue sequence from a database comprising a plurality of known plant
 CC sequences comprising inputting sequence information selected from one of
 CC 464 fully defined sequences given in the specification. The isolated or
 CC recombinant polynucleotide is used for producing a plant having a
 CC modified trait, the method comprising selecting a polynucleotide that
 CC encodes a polypeptide or an antisense nucleic acid, inserting the
 CC polynucleotide or antisense nucleic acid into an expression vector,
 CC introducing the vector into a plant or a cell of a plant to overexpress
 CC the polypeptide or antisense nucleic acid, thereby producing a modified
 CC plant, and selecting for a modified trait (e.g. increased
 CC production of agriculturally useful proteins or metabolic chemicals,
 CC pest tolerance, environmental stress response (e.g. drought), microbial
 CC disease resistance, herbicide resistance, seed and fruit yield, growth
 CC rate, leaf and flower senescence and many other traits listed in the
 CC specification). The present sequence is one of the 232 proteins which are
 CC A. thaliana transcription factors.
 CC
 CC
 CC Sequence 769 AA;
 XX
 SO
 Query Match 3.3%; Score 99; DB 23; Length 769;
 Best Local Similarity 21.8%; Pred. No. 2.9; Indels 78; Gaps 19;
 Matches 85; Conservative 54; Mismatches 173; Indels 78; Gaps 19;
 Oy 66 PSDNRIMGFAEARVLD-----AGVFP-----NTNTNSDKILAGLISGFDE---- 108
 Db 24 PSSNQTLGLANGFYLDLDFSSLDPEEAVPSQNNNNNNINNKAVAGDILLSSSDDADFS 83
 Oy 109 DSCLSRYQSVHNRKPPSPSSYLIS-KLRNYEK-LHKRCGPGTESKAKALKQDQEHID 166
 Db 84 DSVLKYSIQVLMEDMEKPCMFADALALAAREKSLYEALG---EKYSSSSASVSHPE 140
 Oy 167 -----GDGECKYVWISFSGILGRILSLASFVFLYALLTDRVLVDRCKMDLDFCEPFL 220
 Db 141 RLASDSPDGSCS-----GGAESDYAS-----TTTTSSDSHWVYDGLENRP-- 181
 Oy 221 GMSWL---LPLDFPMTDQFDGLNQESSRCYGVAKNOVIDTEGTLSHLYLHVHDGHD 277
 Db 182 --SWLHTPMPNSF---VFQSTSRNSVNTGGGGGCGNSAVYGGSGDGLVSNMFKD--DEL 233
 Oy 278 KMFCEGDDQ---TITGKVPWLYKTPNYFVPSLWLIPGEDELNKLKLF---PQKATVPFHL 331

[illegible]

XX	The present invention provides the protein and coding sequences of
CC	proteins from Moraxella catarrhalis. These can be used to produce
CC	vaccines which protect against M. catarrhalis infection, which can cause
CC	otitis media, respiratory infection, sinusitis, and pneumonia. The
CC	present sequence is a protein of the invention.
XX	
SO	Sequence 814 AA;
	Query Match 3.3%; Score 99; DB 23; Length 814;
	Best Local Similarity 18.1%; Pred. No. 3.2;
	Matches 79; Conservative 49; Mismatches 149; Indels 160; Gaps 16;
OY	153 YKKALKQDQHIDGDGECKYYVWISFSGLNRIILSLASVFLYALLTDRLVLVDKDKMD 212
Db	238 YKDYLKHLPDSHAKSNAGTLGVSW- - -GNQGFGLASV- - -SLRDKYGLPNHSHEYE 289
OY	213 DLFCPPFLGMSWLPLDLPRMTDQELGNDGS- - -SKCYGYWKKNQYIDTEGILSHLY 266
Db	290 E- -C- - - - -SVHGISOALQKYRPYLRLPYLFPELMENDLDFPDNAGLECH 329
OY	267 LHLVHDYGDHDKMFCESGQPTFGIVPMLIYKTDMYVPWSLLIGFDDELTKLRPOKAT 326
Db	330 THDDHH- EHDARDHEDHEDHKRPWIDLKMKRIDY- - -OGQINAPFAGIDK 379
OY	327 VFNHILGRYLFHPITNOVGWGLVTREAYLSHADEKIGIQVRVEDDPGPQHVMDOISSCT 386
Db	380 IRASNGKVYDH- - - - -HDELDGEKTSFFDNQAN- - - - - 408
OY	387 QKEKLIPREVDTLVERSRHVNTPKKAVALVTS- - -LANGAENLKSMWEYPTSTGELIGVH 444
Db	409 - - - - -VWRLEASHPTPIHTPMCGSFSGVFEGYLTSSKNS- - - - -GLV 443
OY	445 OPSOE- -GYQTQEKEMHNKGKALAEWYLLSLTQNIVTSAMSTEGYVAQGIGLKPWILYP 502
Db	444 PRPYDGMKODTONTLHNKK- - - - -TKTGSVTFEE- - - - -YKP 477
OY	503 ENRRTPDPSGC- - - - -GRAMSMEPCFSPFPYDCAKATGIDTGT 540
Db	478 NDKLVLDAARIEKQTIIMDYDKDALIYOSLNLGLTAHER- - - - -DIRKRLLDSC 529
OY	541 LVPHVRHCEDIISWGKL 557
Db	530 LNPKKQTARSYAVGTHL 546
RESULT 14	
AAB19945	
ID	AAB19945 standard; Protein; 818 AA.
XX	
AC	AAB19945;
XX	
DT	19-MAR-2001 (first entry)
XX	
DE	Moraxella catarrhalis BASB107.
XX	
KW	BASB107: infection; pneumonia; otitis media; sinusitis;
KW	therapy; diagnosis; vaccine; genetic immunization; antibacterial;
XX	antibiotic; receptor.
OS	Moraxella catarrhalis.
PX	
MO	WO200071724-A2.
XX	
PD	30-NOV-2000.
PF	
FE	18-MAY-2000; 2000WO-EP04618.
PR	24-MAY-1999; 99GB-0012038.
PR	24-MAY-1999; 99GB-0012040.
PR	01-JUN-1999; 99GB-0012674.
PR	01-JUN-1999; 99GB-0012705.
PR	02-JUN-1999; 99GB-0012838.

[illegible][illegible]

OY 368 FEDEBPFQHVMDIISCTQKEKLLPEVDTLVERSRHVNTPKHKAVALYTSLNAGY----- 422
Db 332 -----NYLIRHPSCVNVSKMNAVICSGTYAQYVQTM 363
OY 423 -AENIKSMY--EYPTSTGEITGVHQ---PSQEGYQOTEK--KMH-NGKA--LAEMYL 470
Db 364 STONLSMTITRDEYPSNPMVLKINGKAAFPQYQPVVMLEKGYTIHNGPAPRTFELYLV 423
OY 471 SLTDN-----LYTSAMSTFGYVAGLGGLKFWILYRP 502
Db 424 NENKNDWIRVGLCYPSNTSFOYTFGLORONGSLSKIEYEP 465

Search completed: February 27, 2003, 15:30:40
Job time : 44 secs

